

Measuring our universe from galaxy redshift surveys

Ofer Lahav^{(1),(2)} and Yasushi Suto^{(3),(4)}

⁽¹⁾ Department of Physics and Astronomy, University of London,
Gower Street, London WC1E 6BT, UK

⁽²⁾ Institute of Astronomy, University of Cambridge,
Madingley Road, Cambridge CB3 0HA, UK

⁽³⁾ Department of Physics, The University of Tokyo,
Tokyo 113-0033, Japan

⁽⁴⁾ Research Center for the Early Universe, School of Science,
The University of Tokyo, Tokyo 113-0033, Japan

e-mail: lahav@star.ucl.ac.uk, suto@phys.s.u-tokyo.ac.jp

(last modified: February 2, 2008)

Abstract

Galaxy redshift surveys have achieved significant progress over the last couple of decades. Those surveys tell us in the most straightforward way what our local universe looks like. While the galaxy distribution traces the *bright* side of the universe, detailed quantitative analyses of the data have even revealed the *dark* side of the universe dominated by non-baryonic dark matter as well as more mysterious dark energy (or Einstein's cosmological constant). We describe several methodologies of using galaxy redshift surveys as cosmological probes, and then summarize the recent results from the existing surveys. Finally we present our views on the future of redshift surveys in the era of Precision Cosmology.

Contents

1	Introduction	4
2	Clustering in the expanding universe	6
2.1	The cosmological principle	6
2.2	From the Einstein equation to the Friedmann equation	8
2.3	Expansion law and age of the universe	11
2.4	Einstein's static model and Lemaître's model	12
2.5	Vacuum energy as an effective cosmological constant	14
2.6	Gravitational instability	14
2.7	Linear growth rate of the density fluctuation	16
3	Statistics of cosmological density fluctuations	19
3.1	Gaussian random field	19
3.2	Log-normal distribution	21
3.3	Higher-order correlation functions	23
3.4	Genus statistics	24
3.5	Minkowski functionals	27
4	Galaxy biasing	29
4.1	Concepts and definitions of biasing	29
4.2	Modeling biasing	29
4.3	Density peaks and dark matter halos as toy models for galaxy biasing	31
4.4	Biasing of galaxies in cosmological hydrodynamic simulations . .	36
4.5	Halo occupation function approach for galaxy biasing	42
5	Relativistic effects observable in clustering at high redshifts	44
5.1	Cosmological lightcone effect on the two-point correlation functions	44
5.2	Evaluating two-point correlation functions from N-body simula- tion data	47
5.3	Cosmological redshift-space distortion	50
5.4	Two-Point Clustering Statistics on a Light-Cone in Cosmological Redshift Space	53
6	Recent results from 2dF and SDSS	58
6.1	The latest galaxy redshift surveys	58
6.1.1	The 2dF galaxy redshift survey	58
6.1.2	The SDSS galaxy redshift survey	59
6.1.3	The 6dF galaxy redshift survey	62
6.1.4	The DEEP galaxy redshift survey	62
6.1.5	The VIRMOS galaxy redshift survey	62
6.2	Cosmological parameters from 2dFGRS	63

6.2.1	The Power spectrum of 2dF Galaxies on large scales . . .	63
6.2.2	An upper limit on neutrino masses	63
6.2.3	Combining 2dFGRS & CMB	63
6.2.4	Redshift-space distortion	65
6.2.5	The bi-spectrum and higher moments	65
6.3	Luminosity and spectral-type dependence of galaxy clustering . .	66
6.3.1	2dFGRS: Clustering per luminosity and spectral type . .	66
6.3.2	SDSS: Two-point correlation functions per luminosity and color	69
6.3.3	SDSS: Three-point correlation functions and the nonlinear biasing of galaxies per luminosity and color	70
6.4	Topology of the universe: analysis of SDSS galaxies in terms of Minkowski functionals	72
6.5	Other statistical measures	75
7	Discussion	76
8	Acknowledgements	78

1 Introduction

Nowadays the exploration of the universe can be performed by a variety of observational probes and methods over a wide range of the wavelengths; temperature anisotropy map of the cosmic microwave background (CMB), the Hubble diagrams (i.e., redshift–magnitude relations) of nearby galaxies and distant Type Ia supernovae, wide-field photometric and spectroscopic surveys of galaxies, power spectrum and abundances of galaxy clusters in optical and X-ray bands combined with the radio observation through the Sunyaev–Zel’dovich effect, deep surveys of galaxies in submm, infrared, and optical bands, quasar surveys in radio and optical, strong and weak lensing of distant galaxies and quasars, high-energy cosmic rays and so on. Undoubtedly gamma-rays, neutrinos, gravitational radiations will join the above already crowded list.

Among those, optical galaxy redshift surveys are the most classical. Indeed one may phrase that the modern observational cosmology started with a sort of a galaxy redshift survey by Edwin Hubble. Still galaxy redshift surveys are of vital importance in cosmology in the 21st century for various reasons:

Redshift surveys have unprecedented quantity and quality:

The numbers of galaxies and quasars in the spectroscopic sample of Two Degree Field (2dF: see §6.1.1) are $\sim 250,000$ and $\sim 30,000$, and will reach $\sim 800,000$ and $100,000$ upon completion of the on-going Sloan Digital Sky Survey (SDSS: see §6.1.2). These unprecedented numbers of the objects as well as the homogeneous selection criteria enable the precise statistical analysis of their distribution.

The universe at $z \approx 1000$ is well specified:

The first-year WMAP (*Wilkinson Microwave Anisotropy Probe*) data [6] among others have established a set of cosmological parameters. This may be taken as the *initial condition* of the universe from the point-of-view of the structure evolution toward $z = 0$. In a sense, the origin of the universe at $z \approx 1000$ and the evolution of the universe after the epoch are now equally important, but well separable questions that particle and observational cosmologists focus on, respectively.

Gravitational growth of dark matter component is well understood:

In addition, extensive numerical simulations of structure formation in the universe has significantly advanced our understanding of the gravitational evolution of dark matter component in the standard gravitational instability picture. In fact, we even have very accurate and useful analytic formulae to describe the evolution deep in its nonlinear regime. Thus we can now directly address the evolution of *visible objects* from the analysis of their redshift surveys separately from the nonlinear growth of the underlying dark matter gravitational potentials.

Formation and evolution of galaxies:

In the era of precision cosmology among others, scientific goals of researches out of galaxy redshift surveys are gradually shifting from inferring a set of values of cosmological parameters using galaxy as their probes to understanding the origin and evolution of galaxy distribution given a set of parameters accurately determined by the other probes like CMB and supernovae.

With the above in mind, we will attempt to summarize what we have learned so far from galaxy redshift surveys and then to describe what will be done with future data. The review is organized as follows. We first present a brief overview of the Friedmann model and gravitational instability theory in §2. Then we describe the non-Gaussian nature of density fluctuations generated by the nonlinear gravitational evolution of the primordial Gaussian field (§3). Next we discuss the spatial biasing of galaxies relative to the underlying dark matter distribution in §4. Our understanding of biasing is still far from complete, and its description is necessarily empirical and very approximate. Nevertheless this is one of the most important ingredients for proper interpretation of galaxy redshift surveys. Section 5 introduces general relativistic effects which become important especially for galaxies at high redshifts. We present the latest results from the two currently largest galaxy redshift surveys, 2dF (Two Degree Field) and SDSS (Sloan Digital Sky Survey), in §6. Finally §7 is devoted to summary of the present knowledge of our universe and our personal view of the future direction of cosmological researches in the new millennium.

2 Clustering in the expanding universe

2.1 The cosmological principle

Our current universe exhibits a wealth of nonlinear structures, but the zero-th order description of our universe is based on the assumption that the universe is homogeneous and isotropic smoothed over sufficiently large scales. This statement is usually referred to as *the cosmological principle*. In fact, the cosmological principle was first adopted when observational cosmology was in its infancy; it was then little more than a conjecture, embodying 'Occam's razor' for the simplest possible model.

Rudnicki (1995) [73] summarized various forms of cosmological principles in modern-day language, which were stated over different periods in human history based on philosophical and aesthetic considerations rather than on fundamental physical laws.

(i) The ancient Indian cosmological principle:

The Universe is infinite in space and time and is infinitely heterogeneous.

(ii) The ancient Greek cosmological principle:

Our Earth is the natural center of the Universe.

(iii) The Copernican cosmological principle:

The Universe as observed from any planet looks much the same.

(iv) The (generalized) cosmological principle:

The Universe is (roughly) homogeneous and isotropic.

(v) The perfect cosmological principle:

The Universe is (roughly) homogeneous in space and time, and is isotropic in space.

(vi) The anthropic principle:

A human being, as he/she is, can exist only in the Universe as it is.

We note that the ancient Indian principle may be viewed as a sort of 'fractal model'. The perfect cosmological principle led to the steady state model, which although more symmetric than the (generalized) cosmological principle, was rejected on observational grounds. The anthropic principle is becoming popular again, e.g. in 'explaining' the non-zero value of the cosmological constant.

Like with any other idea about the physical world, we cannot prove a model, but only falsify it. Proving the homogeneity of the Universe is in particular difficult as we observe the Universe from one point in space, and we can only deduce isotropy indirectly. The practical methodology we adopt is to assume homogeneity and to assess the level of fluctuations relative to the mean, and

hence to test for consistency with the underlying hypothesis. If the assumption of homogeneity turns out to be wrong, then there are numerous possibilities for inhomogeneous models, and each of them must be tested against the observations.

For that purpose, one needs observational data with good quality and quantity extending up to high redshifts. Let us mention some of those.

The CMB Fluctuations Ehlers, Garen and Sachs (1968) [20] showed that by combining the CMB isotropy with the Copernican principle one can deduce homogeneity. More formally their theorem (based on the Liouville theorem) states that “If the fundamental observers in a dust space-time see an isotropic radiation field, then the space-time is locally given by the Friedman-Robertson-Walker (FRW) metric”. The COBE (COsmic Background Explorer) measurements of temperature fluctuations $\Delta T/T = 10^{-5}$ on scales of 10° give via the Sachs Wolfe effect ($\Delta T/T = \frac{1}{3}\Delta\phi/c^2$) and Poisson equation rms density fluctuations of $\frac{\delta\rho}{\rho} \sim 10^{-4}$ on $1000 h^{-1} \text{ Mpc}$ (e.g. [96]), which implies that the deviations from a smooth Universe are tiny.

Galaxy Redshift Surveys The distribution of galaxies in local redshift surveys is highly clumpy, with the Supergalactic Plane seen in full glory. However, deeper surveys like 2dF and SDSS (see §6) show that the fluctuations decline as the length-scales increase. Peebles (1993) [69] has shown that the angular correlation functions for the Lick and APM (Automatic Plate Measuring) surveys scale with magnitude as expected in a universe which approaches homogeneity on large scales. While redshift surveys can provide interesting estimates of the fluctuations on intermediate scales (e.g. [72]), the problems of biasing, evolution and K -correction, would limit the ability of those redshift surveys to ‘prove’ the Cosmological Principle. Despite these worries the measurement of the power spectrum of galaxies derived on the assumption of an underlying FRW metric shows good agreement with the Λ -CDM (cold dark matter) model.

Radio Sources Radio sources in surveys have typical median redshift $\bar{z} \sim 1$, and hence are useful probes of clustering at high redshift. Unfortunately, it is difficult to obtain distance information from these surveys: the radio luminosity function is very broad, and it is difficult to measure optical redshifts of distant radio sources. Earlier studies claimed that the distribution of radio sources supports the ‘Cosmological Principle’. However, the wide range in intrinsic luminosities of radio sources would dilute any clustering when projected on the sky. Recent analyses of new deep radio surveys suggest that radio sources are actually clustered at least as strongly as local optical galaxies. Nevertheless, on the very large scales the distribution of radio sources seems nearly isotropic.

The XRB The X-ray Background (XRB) is likely to be due to sources at high redshift. The XRB sources are probably located at redshift $z < 5$, making them convenient tracers of the mass distribution on scales intermediate between those in the CMB as probed by COBE, and those probed by optical and IRAS redshift surveys. The interpretation of the results depends somewhat on the nature of the X-ray sources and their evolution. By comparing the predicted multipoles to those observed by HEAO1, Scharf et al. (2000) [75] estimate the amplitude of fluctuations for an assumed shape of the density fluctuations. The observed fluctuations in the XRB are roughly as expected from interpolating between the local galaxy surveys and the COBE and other CMB experiments. The rms fluctuations $\frac{\delta\rho}{\rho}$ on a scale of $\sim 600h^{-1}\text{Mpc}$ are less than 0.2 %.

Since the (generalized) cosmological principle is now well supported by the above observations, we shall assume below that it holds over scales $l > 100h^{-1}\text{Mpc}$.

The rest of the current section is devoted to a brief review of the homogeneous and isotropic cosmological model. Further details may be easily found in standard cosmology textbooks [93, 62, 69, 64, 10, 63].

The cosmological principle is mathematically paraphrased as that the metric of the universe (in its zero-th order approximation) is given by

$$ds^2 = -dt^2 + a(t)^2 \left(\frac{dx^2}{1 - Kx^2} + x^2 d\theta^2 + x^2 \sin^2 \theta d\phi^2 \right), \quad (1)$$

where x is the comoving coordinate, and we use the units in which the light velocity $c = 1$. The above Robertson – Walker metric is specified by a constant K , the spatial curvature, and a function of time $a(t)$, the scale factor.

The homogeneous and isotropic assumption also implies that the energy momentum tensor of the matter field, $T_{\mu\nu}$, should take the form of the ideal fluid:

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu + p g_{\mu\nu}, \quad (2)$$

where u_μ is the 4-velocity of the matter, ρ is the mean energy density, and p is the mean pressure.

2.2 From the Einstein equation to the Friedmann equation

The next task is to write down the Einstein equation:

$$R_{\mu\nu} - \frac{1}{2}R g_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \quad (3)$$

using equations (1) and (2). In this case one is left with the following two independent equations:

$$\left(\frac{1}{a} \frac{da}{dt} \right)^2 = \frac{8\pi G}{3} \rho - \frac{K}{a^2} + \frac{\Lambda}{3}, \quad (4)$$

$$\frac{1}{a} \frac{d^2 a}{dt^2} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}, \quad (5)$$

for the three independent functions, $a(t)$, $\rho(t)$, and $p(t)$.

Differentiation of equation (4) with respect to t yields

$$\frac{\ddot{a}}{a} = \frac{4\pi G}{3} \left(\dot{\rho} \frac{a}{\dot{a}} + 2\rho \right) + \frac{\Lambda}{3}. \quad (6)$$

Then eliminating \ddot{a} with equation (5), one obtains

$$\dot{\rho} = -3\frac{\dot{a}}{a}(\rho + p). \quad (7)$$

This can be easily interpreted as the 1st law of the thermodynamics $dQ = dU - pdV = d(\rho a^3) - pd(a^3) = 0$ in the present context. Equations (4) and (7) are often used as the two independent basic equations for $a(t)$, instead of equations (4) and (5).

In either case, however, one needs another independent equation to solve for $a(t)$. This is usually given by the equation of state of the form $p = p(\rho)$. In cosmology, the following simple relation:

$$p = w\rho. \quad (8)$$

While the value of w may change as redshift in principle, it is often assumed that w is independent of time just for simplicity. Then substituting this equation of state into equation (7) immediately yields

$$\rho \propto a^{-3(1+w)}. \quad (9)$$

The non-relativistic matter (or dust), ultra-relativistic matter (or radiation), and the cosmological constant correspond to $w = 0$, $1/3$, and -1 , respectively.

If the universe consists of different fluid species with w_i ($i = 1, \dots, N$), equation (9) still holds independently as long as they do not interact with each other. If one denotes the present energy density of the i -th component by $\rho_{i,0}$, then the total energy density of the universe at the epoch corresponding to the scale factor of $a(t)$ is given by

$$\rho = \sum_{i=1}^N \frac{\rho_{i,0}}{a^{3(1+w_i)}}, \quad (10)$$

where the present value of the scale factor, a_0 , is set to be unity without loss of generality. Thus, equation (4) becomes

$$\left(\frac{1}{a} \frac{da}{dt} \right)^2 = \frac{8\pi G}{3} \sum_{i=1}^N \frac{\rho_{i,0}}{a^{3(1+w_i)}} - \frac{K}{a^2} + \frac{\Lambda}{3}. \quad (11)$$

Note that those components with $w_i = -1$ may be equivalent to the conventional cosmological constant Λ at this level, although they may exhibit spatial variation unlike Λ .

Evaluating equation (11) at the present epoch, one finds

$$H_0^2 = \frac{8\pi G}{3} \sum_{i=1}^N \rho_{i,0} - K + \frac{\Lambda}{3}, \quad (12)$$

where H_0 is the Hubble constant at the present epoch. The above equation is usually rewritten as follows:

$$K \equiv H_0^2 \Omega_K = H_0^2 \left(\sum_{i=1}^N \Omega_{i,0} + \Omega_\Lambda - 1 \right), \quad (13)$$

where the density parameter for the i -th component is defined as

$$\Omega_{i,0} \equiv \frac{8\pi G}{3H_0^2} \rho_{i,0}, \quad (14)$$

and similarly the dimensionless cosmological constant is

$$\Omega_\Lambda \equiv \frac{\Lambda}{3H_0^2}. \quad (15)$$

Incidentally equation (13) clearly illustrates the Mach principle in the sense that the space curvature is simply determined by the amount of matter components in the universe. In particular, the flat universe ($K = 0$) implies that the sum of the density parameters is unity:

$$\sum_{i=1}^N \Omega_{i,0} + \Omega_\Lambda = 1. \quad (16)$$

Finally the cosmic expansion is described by

$$\left(\frac{da}{dt} \right)^2 = H_0^2 \left(\sum_{i=1}^N \frac{\Omega_{i,0}}{a^{1+3w_i}} + 1 - \sum_{i=1}^N \Omega_{i,0} - \Omega_\Lambda + \Omega_\Lambda a^2 \right). \quad (17)$$

As will be shown below, the present universe is supposed to be dominated by non-relativistic matter (baryons and collisionless dark matter) and the cosmological constant. So in the present review, we approximate equation (17) as

$$\left(\frac{da}{dt} \right)^2 = H_0^2 \left(\frac{\Omega_m}{a} + 1 - \Omega_m - \Omega_\Lambda + \Omega_\Lambda a^2 \right) \quad (18)$$

unless otherwise stated.

2.3 Expansion law and age of the universe

Equation (18) has the following analytic solutions in several simple but practically important cases.

- (a) Einstein – de Sitter model ($\Omega_m = 1, \Omega_\Lambda = 0$)

$$a(t) = \left(\frac{t}{t_0}\right)^{2/3}, \quad t_0 = \frac{2}{3H_0} \quad (19)$$

- (b) Open model with vanishing cosmological constant ($\Omega_m < 1, \Omega_\Lambda = 0$)

$$a = \frac{\Omega_m}{2(1 - \Omega_m)} (\cosh \theta - 1), \quad H_0 t = \frac{\Omega_m}{2(1 - \Omega_m)^{3/2}} (\sinh \theta - \theta) \quad (20)$$

$$H_0 t_0 = \frac{1}{1 - \Omega_m} - \frac{\Omega_m}{2(1 - \Omega_m)^{3/2}} \ln \frac{2 - \Omega_m + 2\sqrt{1 - \Omega_m}}{\Omega_m} \quad (21)$$

- (c) Spatially-flat model with cosmological constant ($\Omega_m < 1, \Omega_\Lambda = 1 - \Omega_m$)

$$a(t) = \left(\frac{\Omega_m}{1 - \Omega_m}\right)^{1/3} \left[\sinh \frac{3\sqrt{1 - \Omega_m}}{2} H_0 t \right]^{2/3}, \quad (22)$$

$$H_0 t_0 = \frac{1}{3\sqrt{1 - \Omega_m}} \ln \frac{2 - \Omega_m + 2\sqrt{1 - \Omega_m}}{\Omega_m} \quad (23)$$

In the above, t_0 denotes the present age of the universe:

$$t_0 = \frac{1}{H_0} \int_0^1 \frac{x dx}{\sqrt{\Omega_m x + (1 - \Omega_m - \Omega_\Lambda)x^2 + \Omega_\Lambda x^4}}, \quad (24)$$

and we adopt the initial condition that $a = 0$ at $t = 0$. The expression clearly indicates that t_0 increases as Ω_m decreases and/or Ω_Λ increases. Figure 1 plots the scale factor as a function of $H_0(t - t_0)$, and Table 1 summarizes the age of the universe.

Table 1: The present age of the universe $[(h/0.7)^{-1} \text{ Gyr}]$.

Ω_m	Open model ($\Omega_\Lambda = 0$)	Spatially-flat model ($\Omega_\Lambda = 1 - \Omega_m$)
1.0	9.3	9.3
0.5	10.5	11.6
0.3	11.3	13.5
0.1	12.5	17.8
0.01	13.9	28.0

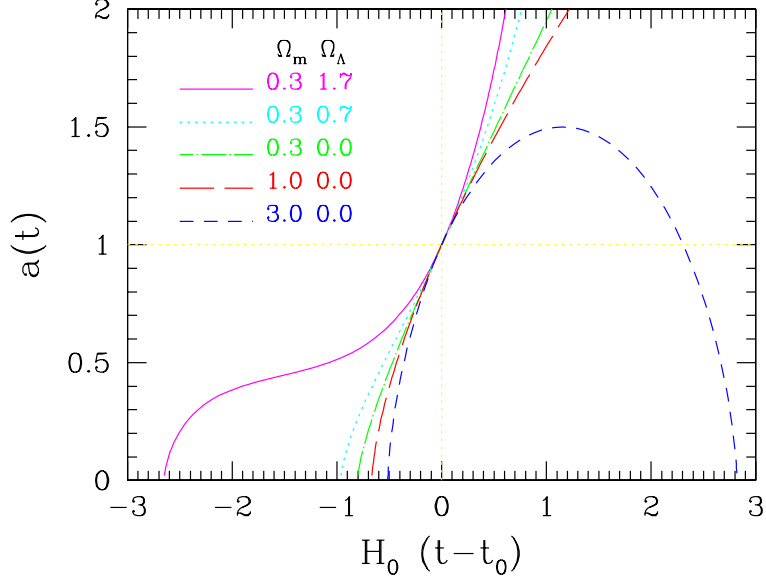


Figure 1: Evolution of the cosmic scale factor as a function of $H_0(t - t_0)$. The present value of the scale factor a_0 is set to unity; solid line: $(\Omega_m, \Omega_\Lambda) = (0.3, 1.7)$, dotted line: $(0.3, 0.7)$, dot-dashed line: $(0.3, 0.0)$, long dashed line: $(1.0, 0.0)$, short dashed line: $(3.0, 0.0)$.

2.4 Einstein's static model and Lemaître's model

So far we have shown that solutions of the Einstein equation are *dynamical* in general, i.e., the scale factor a is time-dependent. As a digression, let us examine why Einstein once introduced the Λ -term to obtain a static cosmological solution. This is mainly important for historical reasons, but also is interesting to observe how the operationally identical parameter (the Λ -term, the cosmological constant, the vacuum energy, the dark energy) shows up in completely different contexts in the course of the development of cosmological physics.

Consider first the case of $\Lambda = 0$ in equations (4) and (5). Clearly the necessary and sufficient condition that the equations admit the solution of $a = \text{const.}$ is given by

$$\rho = -3p = \frac{3K}{8\pi G a^2}. \quad (25)$$

Namely, any static model requires that the universe is dominated by matter with either negative pressure or negative density. This is physically unacceptable as long as one considers *normal* matter in the standard model of particle physics.

If $\Lambda \neq 0$ on the other hand, the condition for the static solution is

$$K = 4\pi G(\rho + p)a^2, \quad \Lambda = 4\pi G(\rho + 3p), \quad (26)$$

yielding

$$\rho = \frac{1}{8\pi G} \left(\frac{3K}{a^2} - \Lambda \right), \quad p = \frac{1}{8\pi G} \left(\Lambda - \frac{K}{a^2} \right). \quad (27)$$

Thus both ρ and p can be positive if

$$\frac{K}{a^2} \leq \Lambda \leq \frac{3K}{a^2}. \quad (28)$$

In particular, if $p = 0$,

$$\rho = \frac{\Lambda}{4\pi G}, \quad K = \Lambda a^2. \quad (29)$$

This represents the closed universe (with positive spatial curvature), and corresponds to Einstein's static model.

The above static model is a special case of Lemaître's universe model with $\Lambda > 0$ and $K > 0$. For simplicity, let us assume that the universe is dominated by non-relativistic matter with negligible pressure, and consider the behavior of Lemaître's model. First define the values of the density and the scale factor corresponding to Einstein's static model:

$$\rho_E = \frac{\Lambda}{4\pi G}, \quad K = \Lambda a_E^2. \quad (30)$$

In order to study the stability of the model around the static model, consider a model in which the density at $a = a_E$ is a factor of $\alpha (> 1)$ larger than ρ_E . Then

$$\rho = \alpha \rho_E \left(\frac{a_E}{a} \right)^3 = \frac{\alpha \Lambda}{4\pi G} \left(\frac{a_E}{a} \right)^3, \quad (31)$$

and equations (4) and (5) reduce to

$$\dot{a}^2 = \frac{\Lambda}{3} \left(\frac{2\alpha a_E^3}{a} - 3a_E^2 + a^2 \right), \quad (32)$$

$$\frac{\ddot{a}}{a} = \frac{\Lambda}{3} \left[1 - \alpha \left(\frac{a_E}{a} \right)^3 \right]. \quad (33)$$

For the period of $a \ll a_E$, equation (32) indicates that $a \propto t^{2/3}$ and the universe is decelerating ($\ddot{a} < 0$). When a reaches $\alpha^{1/3} a_E$, \dot{a}^2 takes the minimum value $\Lambda a_E^2 (\alpha^{2/3} - 1)$ and the universe becomes accelerating ($\ddot{a} > 0$). Finally the universe approaches the exponential expansion or de Sitter model: $a \propto \exp(t\sqrt{\Lambda/3})$. If α becomes closer to unity, the minimum value reaches zero

and the expansion of the universe is effectively frozen. This phase is called the coasting period, and the case with $\alpha = 1$ corresponds to Einstein's static model in which the coasting period continues forever. A similar consideration for $\alpha < 1$ indicates that the universe starts collapsing ($\dot{a}^2 = 0$) before $\ddot{a} = 0$. Thus the behavior of Lemaître's model is crucially different if α is larger or smaller than unity. This suggests that Einstein's static model ($\alpha = 1$) is unstable.

2.5 Vacuum energy as an effective cosmological constant

So far we discussed the cosmological constant introduced in the l.h.s. of the Einstein equation. Formally one can move the Λ -term to the r.h.s. by assigning

$$\rho_\Lambda = \frac{\Lambda}{8\pi G}, \quad p_\Lambda = -\frac{\Lambda}{8\pi G}. \quad (34)$$

This *effective* matter field, however, should satisfy an equation of state of $p = -\rho$. Actually the following example presents a specific example for an *effective* cosmological constant. Consider a real scalar field whose Lagrangian density is given by

$$\mathcal{L} = \frac{1}{2}g_{\mu\nu}\partial^\mu\phi\partial^\nu\phi - V(\phi). \quad (35)$$

Its energy-momentum tensor is

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - \mathcal{L}g_{\mu\nu}, \quad (36)$$

and if the field is spatially homogeneous, its energy density and pressure are

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi), \quad p_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi). \quad (37)$$

Clearly if the evolution of the field is negligible, i.e., $\dot{\phi}^2 \ll V(\phi)$, $p_\phi \approx -\rho_\phi$ and the field acts as a cosmological constant. Of course this model is one of the simplest examples, and one may play with much more complicated models if needed.

If the Λ -term is introduced in the l.h.s., it should be constant to satisfy the energy-momentum conservation $T_{\mu\nu}{}^{;\nu} = 0$. Once it is regarded as a sort of matter field in the r.h.s., however, it does not have to be constant. In fact, the above example shows that the equation of state for the field has $w = -1$ only in special cases. This is why recent literature refers to the field as *dark energy* instead of the cosmological constant.

2.6 Gravitational instability

We have presented the zero-th order description of the universe neglecting the inhomogeneity or spatial variation of matter inside. Now we are in a position to consider the evolution of matter in the universe. For simplicity we focus on

the non-relativistic regime where Newtonian approximation is valid. Then the basic equations for the self-gravitating fluid are given by the continuity equation, Euler's equation and the Poisson equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \quad (38)$$

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\frac{1}{\rho} \nabla p - \nabla \Phi, \quad (39)$$

$$\nabla^2 \Phi = 4\pi G \rho. \quad (40)$$

We would like to rewrite those equations in the comoving frame. For this purpose, we introduce the position \mathbf{x} in the comoving coordinate, the peculiar velocity, \mathbf{v} , density fluctuations $\delta(t, \mathbf{x})$, and the gravitational potential $\phi(t, \mathbf{x})$ which are defined respectively as

$$\mathbf{x} = \frac{\mathbf{r}}{a(t)}, \quad (41)$$

$$\mathbf{v} = a(t) \dot{\mathbf{x}}, \quad (42)$$

$$\delta(t, \mathbf{x}) = \frac{\rho(t, \mathbf{x})}{\bar{\rho}(t)} - 1, \quad (43)$$

$$\phi(t, \mathbf{x}) = \Phi + \frac{1}{2} a \ddot{a} |\mathbf{x}|^2. \quad (44)$$

Then equations (38) to (40) reduce to

$$\dot{\delta} + \frac{1}{a} \nabla \cdot [(1 + \delta) \mathbf{v}] = 0, \quad (45)$$

$$\dot{\mathbf{v}} + \frac{1}{a} (\mathbf{v} \cdot \nabla) \mathbf{v} + \frac{\dot{a}}{a} \mathbf{v} = -\frac{1}{\rho a} \nabla p - \frac{1}{a} \nabla \phi, \quad (46)$$

$$\nabla^2 \phi = 4\pi G \bar{\rho} a^2 \delta, \quad (47)$$

where the dot and ∇ in the above equations are the time derivative for a given \mathbf{x} and the spatial derivative with respect to \mathbf{x} , i.e., defined in the comoving coordinate (while those in eqs.[38] to [40] are defined in the proper coordinate).

A standard picture of the cosmic structure formation assumes that the initially tiny amplitude of density fluctuation grow according to equations (45) to (47). Also the universe smoothed over large scales approaches a homogeneous model. Thus at early epochs and/or on large scales, the nonlinear effect is small and one can linearize those equations with respect to δ and \mathbf{v} :

$$\dot{\delta} + \frac{1}{a} \nabla \cdot \mathbf{v} = 0, \quad (48)$$

$$\dot{\mathbf{v}} + \frac{\dot{a}}{a} \mathbf{v} = -\frac{c_s^2}{a} \nabla \delta - \frac{1}{a} \nabla \phi, \quad (49)$$

$$\nabla^2 \phi = 4\pi G \bar{\rho} a^2 \delta, \quad (50)$$

where $c_s^2 \equiv (\partial p / \partial \rho)$ is the sound velocity.

As usual, we transform the above equations in \mathbf{k} space using

$$\delta_{\mathbf{k}}(t) \equiv \frac{1}{V} \int \delta(t, \mathbf{x}) \exp(i\mathbf{k} \cdot \mathbf{x}) d\mathbf{x}. \quad (51)$$

Then the equation for $\delta_{\mathbf{k}}$ reduces to

$$\ddot{\delta}_{\mathbf{k}} + 2\frac{\dot{a}}{a}\dot{\delta}_{\mathbf{k}} + \left(\frac{c_s^2 k^2}{a^2} - 4\pi G\bar{\rho}\right)\delta_{\mathbf{k}} = 0. \quad (52)$$

If the signature of the third term is positive, $\delta_{\mathbf{k}}$ has an unstable, or, monotonically increasing solution. This condition is equivalent to *the Jeans criterion*:

$$\lambda \equiv \frac{2\pi}{k} > \lambda_J \equiv c_s \sqrt{\frac{\pi}{G\bar{\rho}}}, \quad (53)$$

namely, the wavelength of the fluctuation is larger than the Jeans length λ_J which characterizes the scale that the sound wave can propagate within the dynamical time of the fluctuation $\sqrt{\pi/G\bar{\rho}}$. Below the scale the pressure wave can suppress the gravitational instability, and the fluctuation amplitude oscillates.

2.7 Linear growth rate of the density fluctuation

Most likely our universe is dominated by collisionless dark matter, and thus λ_J is negligibly small. Thus at most scales of cosmological interest, equation (53) is well approximated as

$$\ddot{\delta}_{\mathbf{k}} + 2\frac{\dot{a}}{a}\dot{\delta}_{\mathbf{k}} - 4\pi G\bar{\rho}\delta_{\mathbf{k}} = 0. \quad (54)$$

For a given set of cosmological parameters, one can solve the above equation by substituting the expansion law for $a(t)$ as described in §2.3. Since equation (54) is the second-order differential equation with respect to t , there are two independent solutions; a decaying mode and a growing mode which monotonically decreases and increases as t , respectively. The former mode becomes negligibly small as the universe expands, and thus one is usually interested in the growing mode alone.

More specifically those solutions are explicitly obtained as follows. First note that the l.h.s. of equation (18) is the Hubble parameter at t , $H(t) = \dot{a}/a$:

$$H^2 = H_0^2 \left(\frac{\Omega_m}{a^3} + \frac{1 - \Omega_m - \Omega_\Lambda}{a^2} + \Omega_\Lambda \right) \quad (55)$$

$$= H_0^2 [\Omega_m(1+z)^3 + (1 - \Omega_m - \Omega_\Lambda)(1+z)^2 + \Omega_\Lambda]. \quad (56)$$

The first and second differentiation of equation (55) with respect to t yields

$$2H\dot{H} = H_0^2 \left(-3\frac{\Omega_m}{a^3} - 2\frac{1 - \Omega_m - \Omega_\Lambda}{a^2} \right) H. \quad (57)$$

and

$$\ddot{H} = H_0^2 \left(9 \frac{\Omega_m}{2a^3} + 2 \frac{1 - \Omega_m - \Omega_\Lambda}{a^2} \right) H. \quad (58)$$

Thus the differential equation for H reduces to

$$\ddot{H} + 2H\dot{H} = H_0^2 H \frac{3\Omega_m}{2a^3} = 4\pi G \bar{\rho} H. \quad (59)$$

This coincides with the linear perturbation equation for $\delta_{\mathbf{k}}$ (eq.[54]). Since $H(t)$ is a decreasing function of t , this implies that $H(t)$ is the decaying solution for equation (54). Then the corresponding growing solution $D(t)$ can be obtained according to the standard procedure: subtracting equation (54) from equation (59) yields

$$a^2 \frac{d}{dt} (\dot{D}H - D\dot{H}) + \frac{da^2}{dt} (\dot{D}H - D\dot{H}) = \frac{d}{dt} \left[a^2 H^2 \frac{d}{dt} \left(\frac{D}{H} \right) \right] = 0, \quad (60)$$

and therefore the formal expression for the growing solution in linear theory is

$$D(t) \propto H(t) \int_0^t \frac{dt'}{a^2(t') H^2(t')}. \quad (61)$$

It is often more useful to rewrite $D(t)$ in terms of the redshift z as follows:

$$D(z) = \frac{5\Omega_m H_0^2}{2} H(z) \int_z^\infty \frac{1+z'}{H^3(z')} dz', \quad (62)$$

where the proportional factor is chosen so as to reproduce $D(z) \rightarrow 1/(1+z)$ for $z \rightarrow \infty$. Linear growth rates for the models described in §2.3 are summarized below.

(a) Einstein – de Sitter model ($\Omega_m = 1, \Omega_\Lambda = 0$)

$$D(z) = \frac{1}{1+z} \quad (63)$$

(b) Open model with vanishing cosmological constant ($\Omega_m < 1, \Omega_\Lambda = 0$)

$$D(z) \propto 1 + \frac{3}{x} + 3\sqrt{\frac{1+x}{x^3}} \ln(\sqrt{1+x} - \sqrt{x}), \quad x \equiv \frac{1 - \Omega_m}{\Omega_m(1+z)} \quad (64)$$

(c) Spatially-flat model with cosmological constant ($\Omega_m < 1, \Omega_\Lambda = 1 - \Omega_m$)

$$D(z) \propto \sqrt{1 + \frac{2}{x^3}} \int_0^x \left(\frac{u}{2+u^3} \right)^{3/2} du, \quad x \equiv \frac{2^{1/3}(\Omega_m^{-1} - 1)^{1/3}}{1+z}. \quad (65)$$

For most purposes, the following fitting formulae [65] provide sufficiently accurate approximations.

$$D(z) = \frac{g(z)}{1+z}, \quad (66)$$

$$g(z) = \frac{5\Omega(z)}{2} \frac{1}{\Omega^{4/7}(z) - \lambda(z) + [1 + \Omega(z)/2][1 + \lambda(z)/70]}, \quad (67)$$

where

$$\Omega(z) = \Omega_m(1+z)^3 \left[\frac{H_0}{H(z)} \right]^2 = \frac{\Omega_m(1+z)^3}{\Omega_m(1+z)^3 + (1 - \Omega_m - \Omega_\Lambda)(1+z)^2 + \Omega_\Lambda}, \quad (68)$$

$$\lambda(z) = \Omega_\Lambda \left[\frac{H_0}{H(z)} \right]^2 = \frac{\Omega_\Lambda}{\Omega_m(1+z)^3 + (1 - \Omega_m - \Omega_\Lambda)(1+z)^2 + \Omega_\Lambda}. \quad (69)$$

Note that Ω_m and Ω_Λ refer to the *present* values of the density parameter and the dimensionless cosmological constant, respectively, which will be frequently used in the rest of the review.

Figure 2 shows the comparison of the numerically computed growth rate (thick lines) against the above fitting formulae (thin lines) which are practically indistinguishable.

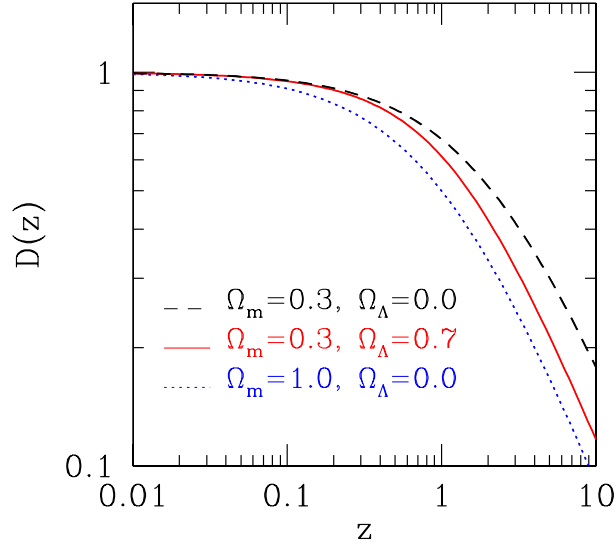


Figure 2: Linear growth rate of density fluctuations.

3 Statistics of cosmological density fluctuations

3.1 Gaussian random field

Consider the density contrast $\delta_i \equiv \delta(\mathbf{x}_i) = \rho(\mathbf{x})/\bar{\rho} - 1$ defined at the comoving position \mathbf{x}_i . The density field is regarded as a stochastic variable, and thus forms a random field. The conventional assumption is that the primordial density field (in its linear regime) is Gaussian, i.e., its m -point joint probability distribution obeys the multi-variate Gaussian:

$$P(\delta_1, \delta_2, \dots, \delta_m) d\delta_1 d\delta_2 \cdots d\delta_m = \frac{1}{\sqrt{(2\pi)^m \det(M)}} \exp \left[- \sum_{i,j=1}^m \frac{1}{2} \delta_i (M^{-1})_{ij} \delta_j \right] d\delta_1 d\delta_2 \cdots d\delta_m, \quad (70)$$

for an arbitrary positive integer m . Here $M_{ij} \equiv \langle \delta_i \delta_j \rangle$ is the covariance matrix, and M^{-1} is its inverse. Since $M_{ij} = \xi(\mathbf{x}_i, \mathbf{x}_j)$, equation (70) implies that the statistical nature of the Gaussian density field is completely specified by the two-point correlation function ξ and its linear combination (including its derivative and integral). For an extensive discussion of the cosmological Gaussian density field, see ref. [5].

The Gaussian nature of the primordial density field is preserved in its linear evolution stage, but this is not the case in nonlinear stage. This is clear even from the definition of the Gaussian distribution: equation (70) formally assumes that the density contrast distributes symmetrically in the range of $-\infty < \delta_i < \infty$, but in the real density field δ_i cannot be less than -1 . This assumption does not make any practical difference as long as the fluctuations are (infinitesimally) small, but is invalid in nonlinear regime where the typical amplitude of the fluctuations exceeds unity.

In describing linear theory of cosmological density fluctuations, the Fourier transform of the spatial density contrast $\delta(\mathbf{x}) \equiv \rho(\mathbf{x})/\langle \bar{\rho} \rangle - 1$ is the most basic variable:

$$\delta_{\mathbf{k}} = \frac{1}{V} \int d\mathbf{x} \delta(\mathbf{x}) \exp(i\mathbf{k} \cdot \mathbf{x}). \quad (71)$$

Since $\delta_{\mathbf{k}}$ is a complex variable, it is decomposed by a set of two real variables, the amplitude $D_{\mathbf{k}}$ and the phase $\phi_{\mathbf{k}}$:

$$\delta_{\mathbf{k}} \equiv D_{\mathbf{k}} \exp(i\phi_{\mathbf{k}}). \quad (72)$$

Then linear perturbation equation reads

$$\ddot{D}_{\mathbf{k}} + 2\frac{\dot{a}}{a}\dot{D}_{\mathbf{k}} - (4\pi G\bar{\rho} + \dot{\phi}^2)D_{\mathbf{k}} = 0, \quad (73)$$

$$\ddot{\phi}_{\mathbf{k}} + 2\left(\frac{\dot{a}}{a} + \frac{\dot{D}_{\mathbf{k}}}{D_{\mathbf{k}}}\right)\dot{\phi}_{\mathbf{k}} = 0. \quad (74)$$

Equation (74) yields $\dot{\phi}(t) \propto a^{-2}(t)D_{\mathbf{k}}^{-2}(t)$, and $\phi(t)$ rapidly converges to a constant value. Thus $D_{\mathbf{k}}$ evolves following the growing solution in linear theory.

The most popular statistic of clustering in the universe is the power spectrum of the density fluctuations:

$$P(t, \mathbf{k}) \equiv \langle D_{\mathbf{k}}(t)^2 \rangle \quad (75)$$

which measures the amplitude of the mode of the wavenumber \mathbf{k} . This is the Fourier transform of the two-point correlation function:

$$\xi(\mathbf{x}, t) = \frac{1}{8\pi^3} \int P(t, \mathbf{k}) \exp(-i\mathbf{k} \cdot \mathbf{x}) d\mathbf{k}. \quad (76)$$

If the density field is globally homogeneous and isotropic (i.e., no preferred position or direction), equation (76) reduces to

$$\xi(x, t) = \frac{1}{2\pi^2} \int_0^\infty P(t, k) \frac{\sin kx}{x} k dk. \quad (77)$$

Since the above expression obtained after the ensemble average, x does not denote an amplitude of the position vector, but a comoving wavelength $2\pi/k$ corresponding to the wavenumber $k = |\mathbf{k}|$. It should be noted that neither the power spectrum or the two-point correlation function contains information for the phase $\phi_{\mathbf{k}}$. Thus in principle two clustering patterns may be completely different even if they have the identical two-point correlation functions. This implies the practical importance to describe the statistics of phases $\phi_{\mathbf{k}}$ in addition to the amplitude $D_{\mathbf{k}}$ of clustering.

In the Gaussian field, however, one can directly show that equation (70) reduces to the probability distribution function of $\phi_{\mathbf{k}}$ and $D_{\mathbf{k}}$ that are explicitly written as

$$P(|\delta_{\mathbf{k}}|, \phi_{\mathbf{k}}) d|\delta_{\mathbf{k}}| d\phi_{\mathbf{k}} = \frac{2|\delta_{\mathbf{k}}|}{P(k)} \exp\left(-\frac{|\delta_{\mathbf{k}}|^2}{P(k)}\right) d|\delta_{\mathbf{k}}| \frac{d\phi_{\mathbf{k}}}{2\pi}. \quad (78)$$

mutually independently of k . The phase distribution is uniform, and thus does not carry information. The above probability distribution function is also derived when the real and imaginary parts of the Fourier components $\delta_{\mathbf{k}}$ are uncorrelated and Gaussian distributed (with the dispersion $P(k)/2$) independently of \mathbf{k} . As is expected, the distribution function (78) is completely fixed if $P(k)$ is specified. This rephrases the previous statement that the Gaussian field is completely specified by the two-point correlation function in real space.

Incidentally the one-point phase distribution turns out to be essentially uniform even in a strongly non-Gaussian field [81, 21]. Thus it is unlikely to extract useful information directly out of it mainly due to the cyclic property of the phase. Very recently, however, Matsubara[52] and Hikage et al.[32] succeeded in detecting a signature of phase correlations in Fourier modes of mass density

fields induced by nonlinear gravitational clustering using the distribution function of the phase sum of the Fourier modes for triangle wavevectors. Several different statistics which carry the phase information have been also proposed in cosmology, including the void probability function [95], the genus statistics [26], the Minkowski functionals [58, 76].

3.2 Log-normal distribution

A probability distribution function (PDF) of the cosmological density fluctuations is the most fundamental statistic characterizing the large-scale structure of the universe. As long as the density fluctuations are in the linear regime, their PDF remains Gaussian. Once they reach the nonlinear stage, however, their PDF significantly deviates from the initial Gaussian shape due to the strong non-linear mode-coupling and the non-locality of the gravitational dynamics. The functional form for the resulting PDFs in nonlinear regimes are not known exactly, and a variety of phenomenological models have been proposed [34, 74, 9, 25].

Kayo et al. (2001) [40] showed that the one-point log-normal PDF:

$$P_{\text{LN}}^{(1)}(\delta) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp \left[-\frac{\{\ln(1+\delta) + \sigma_1^2/2\}^2}{2\sigma_1^2} \right] \frac{1}{1+\delta}. \quad (79)$$

describes very accurately the cosmological density distribution even in the non-linear regime (the rms variance $\sigma_{\text{nl}} \lesssim 4$ and the over-density $\delta \lesssim 100$). The above function is characterized by a single parameter σ_1 which is related to the variance of δ . Since we use δ to represent the density fluctuation field smoothed over R , its variance is computed from its power spectrum P_{nl} explicitly as

$$\sigma_{\text{nl}}^2(R) \equiv \frac{1}{2\pi^2} \int_0^\infty P_{\text{nl}}(k) \tilde{W}^2(kR) k^2 dk. \quad (80)$$

Here we use subscripts “lin” and “nl” to distinguish the variables corresponding to the primordial (linear) and the evolved (nonlinear) density fields, respectively. Then σ_1 depends on the smoothing scale R alone and is given by

$$\sigma_1^2(R) = \ln [1 + \sigma_{\text{nl}}^2(R)]. \quad (81)$$

Given a set of cosmological parameters, one can compute $\sigma_{\text{nl}}(R)$ and thus $\sigma_1(R)$ very accurately using a fitting formula for $P_{\text{nl}}(k)$ (e.g., ref. [65]). In this sense, the above log-normal PDF is completely specified without any free parameter.

Figure 3 plots the one-point PDFs computed from cosmological N-body simulations in SCDM, LCDM and OCDM (for Standard, Lambda and Open CDM) models, respectively [35, 40]. The simulations employ $N = 256^3$ dark matter particles in a periodic comoving cube $(100h^{-1}\text{Mpc})^3$. The density fields are smoothed over Gaussian (*Left panels*) and Top-hat (*Right panels*) windows with

different smoothing lengths; $R = 2h^{-1}\text{Mpc}$, $6h^{-1}\text{Mpc}$ and $18h^{-1}\text{Mpc}$. Solid lines show the log-normal PDFs adopting the value of σ_{nl} directly evaluated from simulations (shown in each panel). The agreement between the log-normal model and the simulation results is quite impressive. A small deviation is noticeable only for $\delta \lesssim -0.5$.

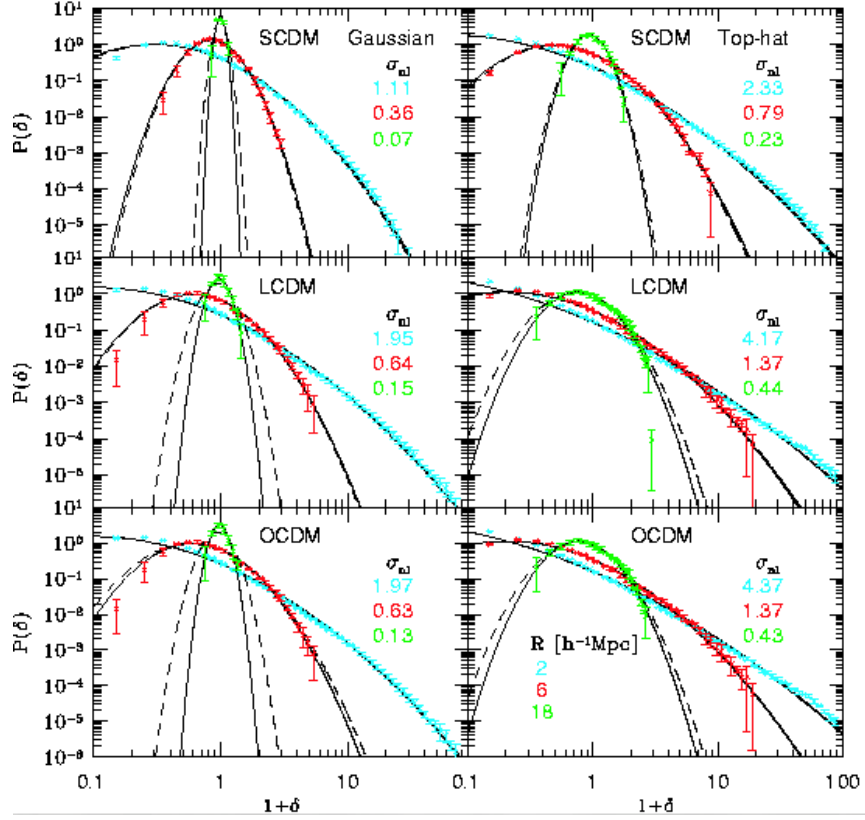


Figure 3: One-point PDFs in CDM models with Gaussian (*left* panels) and top-hat (*right* panels) smoothing windows; $R = 2h^{-1}\text{Mpc}$ (*cyan*), $6h^{-1}\text{Mpc}$ (*red*), and $18h^{-1}\text{Mpc}$ (*green*). The *solid* and *long-dashed* lines represent the log-normal PDF adopting σ_{nl} calculated directly from the simulations and estimated from the nonlinear fitting formula of ref. [65], respectively. (ref. [40])

From an empirical point of view, Hubble (1934) [34] first noted that the galaxy distribution in angular cells on the celestial sphere may be approximated by a log-normal distribution, rather than a Gaussian. Theoretically the above log-normal function may be obtained from the one-to-one mapping between the

linear random-Gaussian and the nonlinear density fields [9]. Define a linear density field g smoothed over R obeying the Gaussian PDF:

$$P_G^{(1)}(g) = \frac{1}{\sqrt{2\pi\sigma_{\text{lin}}^2}} \exp\left(-\frac{g^2}{2\sigma_{\text{lin}}^2}\right), \quad (82)$$

where the variance is computed from its linear power spectrum:

$$\sigma_{\text{lin}}^2(R) \equiv \frac{1}{2\pi^2} \int_0^\infty P_{\text{lin}}(k) \tilde{W}^2(kR) k^2 dk. \quad (83)$$

If one introduces a new field δ from g as

$$1 + \delta = \frac{1}{\sqrt{1 + \sigma_{\text{nl}}^2}} \exp\left\{\frac{g}{\sigma_{\text{lin}}} \sqrt{\ln(1 + \sigma_{\text{nl}}^2)}\right\}, \quad (84)$$

the PDF for δ is simply given by $(dg/d\delta)P_G^{(1)}(g)$ which reduces to equation (79).

At this point, the transformation (84) is nothing but a mathematical procedure to relate the Gaussian and the log-normal functions. Thus there is no physical reason to believe that the new field δ should be regarded as a nonlinear density field evolved from g even in an approximate sense. In fact it is physically unacceptable since the relation, if taken at face value, implies that the nonlinear density field is completely determined by its linear counterpart locally. We know, on the other hand, that the nonlinear gravitational evolution of cosmological density fluctuations proceeds in a quite nonlocal manner, and is sensitive to the surrounding mass distribution. Nevertheless the fact that the log-normal PDF provides a good fit to the simulation data empirically implies that the transformation (84) somehow captures an important aspect of the nonlinear evolution in the real universe.

3.3 Higher-order correlation functions

One of the most direct methods to evaluate the deviation from the Gaussianity is to compute the higher-order correlation functions. Suppose that \mathbf{x}_i now labels the position of the i -th object (galaxy), and then the two-point correlation function $\xi_{12} \equiv \xi(\mathbf{x}_1, \mathbf{x}_2)$ is defined also in terms of the joint probability of the pair of objects located in the volume elements of δV_1 and δV_2 :

$$\delta P_{12} = \bar{n}^2 \delta V_1 \delta V_2 [1 + \xi_{12}], \quad (85)$$

where \bar{n} is the mean number density of the objects. This definition is generalized to three- and four-point correlation functions, $\zeta_{123} \equiv \zeta(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ and $\eta_{1234} \equiv \eta(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$, in a straightforward manner:

$$\begin{aligned} \delta P_{123} &= \bar{n}^3 \delta V_1 \delta V_2 \delta V_3 [1 + \xi_{12} + \xi_{23} + \xi_{31} + \zeta_{123}], \\ \delta P_{1234} &= \bar{n}^4 \delta V_1 \delta V_2 \delta V_3 \delta V_4 [1 + \xi_{12} + \xi_{13} + \xi_{14} + \xi_{23} + \xi_{24} + \xi_{34} \end{aligned} \quad (86)$$

$$\begin{aligned}
& +\zeta_{123} + \zeta_{124} + \zeta_{134} + \zeta_{234} \\
& +\xi_{12}\xi_{34} + \xi_{13}\xi_{24} + \xi_{14}\xi_{23} + \eta_{1234}],
\end{aligned} \tag{87}$$

Apparently ξ_{12} , ζ_{123} , and η_{1234} are symmetric with respect to the change of the indices. Define the following quantities with the same symmetry properties:

$$Z_{123} \equiv \xi_{12}\xi_{23} + \xi_{21}\xi_{13} + \xi_{23}\xi_{31}, \tag{88}$$

$$\begin{aligned}
A_{1234} & \equiv \xi_{12}\xi_{23}\xi_{34} + \xi_{23}\xi_{34}\xi_{41} + \xi_{24}\xi_{41}\xi_{12} + \xi_{13}\xi_{32}\xi_{24} + \xi_{32}\xi_{24}\xi_{41} \\
& + \xi_{24}\xi_{41}\xi_{13} + \xi_{12}\xi_{24}\xi_{43} + \xi_{24}\xi_{43}\xi_{31} + \xi_{31}\xi_{12}\xi_{24} + \xi_{13}\xi_{34}\xi_{42} \\
& + \xi_{34}\xi_{42}\xi_{21} + \xi_{42}\xi_{21}\xi_{13},
\end{aligned} \tag{89}$$

$$B_{1234} \equiv \xi_{12}\xi_{13}\xi_{14} + \xi_{21}\xi_{23}\xi_{24} + \xi_{31}\xi_{32}\xi_{34} + \xi_{41}\xi_{42}\xi_{43}, \tag{90}$$

$$\begin{aligned}
C_{1234} & \equiv \zeta_{123}(\xi_{14} + \xi_{24} + \xi_{34}) + \zeta_{134}(\xi_{12} + \xi_{32} + \xi_{42}) \\
& + \zeta_{124}(\xi_{31} + \xi_{32} + \xi_{34}) + \zeta_{234}(\xi_{12} + \xi_{13} + \xi_{14}).
\end{aligned} \tag{91}$$

Then it is not unreasonable to suspect that the following relations hold

$$\zeta_{123} = Q Z_{123}, \tag{92}$$

$$\eta_{1234} = R_a A_{1234} + R_b B_{1234}, \tag{93}$$

$$\eta_{1234} = R_c C_{1234}, \tag{94}$$

where Q , R_a , R_b , and R_c are constants. In fact, the analysis of the two-dimensional galaxy catalogues [68] revealed

$$Q = 1.29 \pm 0.21 \quad (0.1h^{-1}\text{Mpc} \lesssim r \lesssim 10h^{-1}\text{Mpc}), \tag{95}$$

$$R_a = 2.5 \pm 0.6, \quad R_b = 4.3 \pm 1.2 \quad (0.5h^{-1}\text{Mpc} \lesssim r \lesssim 4h^{-1}\text{Mpc}). \tag{96}$$

The generalization of those relations for N -point correlation functions is suspected to hold generally:

$$\xi_N(\mathbf{r}_1, \dots, \mathbf{r}_N) = \sum_j Q_{N,j} \sum_{(ab)} \prod_{i=1}^{N-1} \xi(r_{ai}), \tag{97}$$

and called *the hierarchical clustering ansatz*. Cosmological N-body simulations approximately support the validity of the above ansatz, but also detect the finite deviation from it [82].

3.4 Genus statistics

A complementary approach to characterize the clustering of the universe beyond the two-point correlation functions is the genus statistics [26]. This is a mathematical measure of the topology of the isodensity surface. For definiteness, consider the density contrast field $\delta(\mathbf{x})$ at the position \mathbf{x} in the survey volume

V_{all} . This may be evaluated, for instance, by taking the ratio of the number of galaxies $N(\mathbf{x}, V_f)$ in the volume V_f centered at \mathbf{x} to its average value $\overline{N}(V_f)$:

$$\delta(\mathbf{x}, V_f) = \frac{N(\mathbf{x}, V_f)}{\overline{N}(V_f)} - 1, \quad \sigma(V_f) = \langle |\delta(\mathbf{x}, V_f)|^2 \rangle^{1/2}, \quad (98)$$

where $\sigma(V_f)$ is its rms value. Consider the isodensity surface parameterized by a value of $\nu \equiv \delta(\mathbf{x}, V_f)/\sigma(V_f)$. Genus is one of the topological numbers characterizing the surface defined as

$$g \equiv -\frac{1}{4\pi} \int \kappa \, dA, \quad (99)$$

where κ is the Gaussian curvature of the isolated surface. The Gauss – Bonnet theorem implies that the value of g is indeed an integer and equal to the number of holes minus 1. This is qualitatively understood as follows; expand an arbitrary two-dimensional surface around a point as

$$z = \frac{1}{2}ax^2 + bxy + \frac{1}{2}cy^2 = \frac{1}{2}\kappa_1x_1^2 + \frac{1}{2}\kappa_2x_2^2. \quad (100)$$

Then the Gaussian curvature of the surface is defined by $\kappa = \kappa_1\kappa_2$. A surface topologically equivalent to a sphere (a torus) has $\kappa = 1$ ($\kappa = 0$), and thus equation (99) yields $g = -1$ ($g = 0$) which coincides with the number of holes minus 1.

In reality, there are many disconnected isodensity surfaces for a given ν , and thus it is more convenient to define the genus density in the survey volume V_{all} using the additivity of the genus:

$$G(\nu) \equiv \frac{1}{V_{\text{all}}} \sum_{i=1}^I g_i = -\frac{1}{4\pi V_{\text{all}}} \sum_{i=1}^I \int_{A_i} \kappa \, dA, \quad (101)$$

where A_i ($i = 1 \sim I$) denote the disconnected isodensity surfaces with the same value of $\nu = \delta(\mathbf{x}, V_f)/\sigma(V_f)$. Interestingly the Gaussian density field has an analytic expression for equation (101):

$$G(\nu) = \frac{1}{4\pi^2} \left(\frac{\langle k^2 \rangle}{3} \right)^{3/2} e^{-\nu^2/2} (1 - \nu^2), \quad (102)$$

where

$$\langle k^2 \rangle \equiv \frac{\int k^2 P(k) \tilde{W}^2(kR) d^3k}{\int P(k) \tilde{W}^2(kR) d^3k} \quad (103)$$

is the moment of k^2 weighted over the power spectrum of fluctuations $P(k)$, and the smoothing function $\tilde{W}^2(kR)$ (e.g., ref. [5]). It should be noted that

in the Gaussian density field the information of the power spectrum shows up only in the proportional constant of equation (102), and its functional form is determined uniquely by the threshold value ν . This ν -dependence reflects the phase information which is ignored in the two-point correlation function and power spectrum. In this sense, genus statistics is a complementary measure of the clustering pattern of universe.

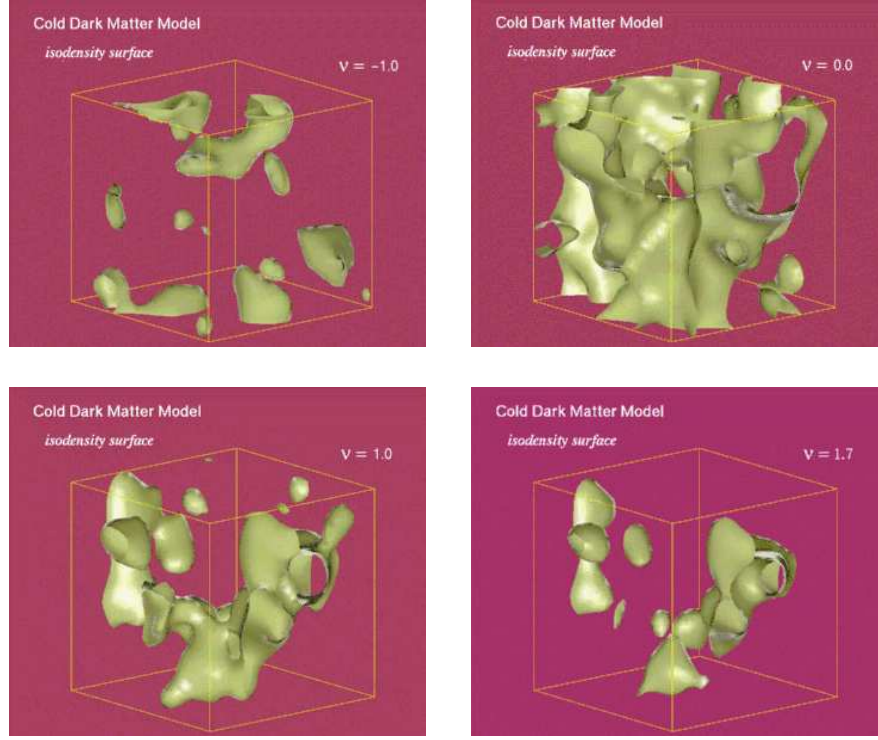


Figure 4: Isodensity surfaces of dark matter distribution from N-body simulation; LCDM in $(100h^{-1}\text{Mpc})^3$ at $\nu = -1.0, 0.0, 1.0$, and 1.7 . (ref. [54])

Even if the primordial density field obeys the Gaussian statistics, the subsequent nonlinear gravitational evolution generates the significant non-Gaussianity. To distinguish the initial non-Gaussianity from that acquired by the nonlinear gravity is of fundamental importance in inferring the initial condition of the universe in a standard gravitational instability picture of structure formation. In a weakly nonlinear regime, Matsubara (1996) derived an analytic expression

for the non-Gaussianity emerging from the primordial Gaussian field [50]:

$$G(\nu) = -\frac{1}{4\pi^2} \left(\frac{\langle k^2 \rangle}{3} \right)^{3/2} e^{-\nu^2/2} \left[H_2(\nu) + \sigma \left(\frac{S}{6} H_5(\nu) + \frac{3T}{2} H_3(\nu) + 3U H_1(\nu) \right) \right], \quad (104)$$

where

$$H_n(\nu) \equiv (-1)^n e^{\nu^2/2} \left(\frac{d}{d\nu} \right)^n e^{-\nu^2/2} \quad (105)$$

are the Hermite polynomials; $H_1 = \nu$, $H_2 = \nu^2 - 1$, $H_3 = \nu^3 - 3\nu$, $H_4 = \nu^4 - 6\nu^2 + 3$, $H_5 = \nu^5 - 10\nu^3 + 15\nu$, ... The three quantities:

$$S = \frac{1}{\sigma^4} \langle \delta^3 \rangle, \quad T = -\frac{1}{2\langle k^2 \rangle \sigma^4} \langle \delta^2 \nabla^2 \delta \rangle, \quad U = -\frac{3}{4\langle k^2 \rangle^2 \sigma^4} \langle \nabla \delta \cdot \nabla \delta \nabla^2 \delta \rangle \quad (106)$$

denote the third-order moments of δ . This expression plays a key role in understanding if the non-Gaussianity in galaxy distribution is ascribed to the primordial departure from the Gaussian statistics.

3.5 Minkowski functionals

In fact, genus is one of the complete sets of $N + 1$ quantities, known as the Minkowski functionals (MFs), which determine the morphological properties of a pattern in N -dimensional space. In the analysis of galaxy redshift survey data, one considers isodensity contours from the three-dimensional density contrast field δ by taking its excursion set F_ν , i.e., the set of all points where the density contrast δ exceeds the threshold level ν as was the case in the case of genus described in the above subsection.

All MFs can be expressed as integrals over the excursion set. While the first MF is simply given by the volume integration of a Heaviside step function Θ normalized to the total volume V_{tot} ,

$$V_0(\nu) = \frac{1}{V_{\text{tot}}} \int_V d^3x \Theta(\nu - \nu(x)), \quad (107)$$

the other MFs, $V_k(k = 1, 2, 3)$, are calculated by the surface integration of the local MFs, v_k^{loc} . The general expression is

$$V_k(\nu) = \frac{1}{V_{\text{tot}}} \int_{\partial F_\nu} d^2S(\mathbf{x}) v_k^{\text{loc}}(\nu, \mathbf{x}), \quad (108)$$

with the local Minkowski Functionals for $k = 1, 2, 3$ given by

$$v_1^{\text{loc}}(\nu, \mathbf{x}) = \frac{1}{6}, \quad (109)$$

$$v_2^{\text{loc}}(\nu, \mathbf{x}) = \frac{1}{6\pi} \left(\frac{1}{R_1} + \frac{1}{R_2} \right), \quad (110)$$

$$v_3^{\text{loc}}(\nu, \mathbf{x}) = \frac{1}{4\pi} \frac{1}{R_1 R_2}, \quad (111)$$

where R_1 and R_2 are the principal radii of curvature of the isodensity surface.

For a 3-D Gaussian random field, the average MFs per unit volume can be expressed analytically as follows:

$$V_0(\nu) = \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^\nu \exp\left(-\frac{x^2}{2}\right) dx, \quad (112)$$

$$V_1(\nu) = \frac{2}{3} \frac{\lambda}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\nu^2\right) \quad (113)$$

$$V_2(\nu) = \frac{2}{3} \frac{\lambda^2}{\sqrt{2\pi}} \nu \exp\left(-\frac{1}{2}\nu^2\right), \quad (114)$$

$$V_3(\nu) = \frac{\lambda^3}{\sqrt{2\pi}} (\nu^2 - 1) \exp\left(-\frac{1}{2}\nu^2\right), \quad (115)$$

where $\lambda = \sqrt{\sigma_1^2/6\pi\sigma^2}$, $\sigma \equiv \langle\delta^2\rangle^{1/2}$, $\sigma_1 \equiv \langle|\nabla\delta|^2\rangle^{1/2}$, and δ is the density contrast.

The above MFs can be indeed interpreted as well-known geometric quantities; the volume fraction $V_0(\nu)$, the total surface area $V_1(\nu)$, the integral mean curvature $V_2(\nu)$, and the integral Gaussian curvature, i.e., the Euler characteristic $V_3(\nu)$. In our current definitions (eqs. [101] and [108], or eqs.[102] and [115]), one can easily show that $V_3(\nu)$ reduces simply to $-G(\nu)$. The MFs were first introduced to cosmological studies by Mecke et al. (1994) [58], and further details may be found in refs. [58, 31]. Analytic expressions of MFs in weakly non-Gaussian fields are derived in ref. [51].

4 Galaxy biasing

4.1 Concepts and definitions of biasing

As discussed above, luminous objects, such as galaxies and quasars, are not direct tracers of the mass in the universe. In fact, the difference of the spatial distribution between luminous objects and dark matter, or the *bias*, has been indicated from a variety of observations. Galaxy “biasing” clearly exists. The fact that galaxies of different types cluster differently (e.g., ref. [16]) implies that not all of them are exact tracers of the underlying mass distribution. See also §6.

In order to confront theoretical model predictions for the *mass* distribution against observational data, one needs a relation of density fields of mass and luminous objects. The biasing of density peaks in a Gaussian random field is well formulated [37, 5] and it provides the first theoretical framework for the origin of galaxy density biasing. In this scheme, the galaxy–galaxy and mass–mass correlation functions are related in the linear regime via

$$\xi_{gg}(r) = b^2 \xi_{mm}(r), \quad (116)$$

where the biasing parameter b is a constant independent of scale r . However, a much more specific linear biasing model is often assumed in common applications, in which the local density fluctuation fields of galaxies and mass are assumed to be deterministically related via the relation

$$\delta_g(\mathbf{x}) = b \delta_m(\mathbf{x}). \quad (117)$$

Note that equation (116) follows from equation (117), but the reverse is not true.

The above deterministic linear biasing is not based on a reasonable physical motivation. If $b > 1$, it must break down in deep voids because values of δ_g below -1 are forbidden by definition. Even in the simple case of no evolution in comoving galaxy number density, the linear biasing relation is not preserved during the course of fluctuation growth. Non-linear biasing, where b varies with δ_m , is inevitable.

Indeed, an analytical model for biasing of halos on the basis of the extended Press-Schechter approximation [59] predicts that the biasing in nonlinear and provides a useful approximation for its behavior as a function of scale, time and mass threshold. N -body simulations provide a more accurate description of the nonlinearity of the halo biasing confirming the validity of the Mo & White model [36, 91].

4.2 Modeling biasing

Biasing is likely to be *stochastic*, not deterministic [15]. An obvious part of this stochasticity can be attributed to the discrete sampling of the density field by

galaxies, i.e. the shot noise. In addition, a statistical, physical scatter in the efficiency of galaxy formation as a function of δ_m is inevitable in any realistic scenario. For example, the random variations in the density on smaller scales is likely to be reflected in the efficiency of galaxy formation. As another example, the local geometry of the background structure, via the deformation tensor, must play a role too. Such ‘hidden variables’ would show up as physical scatter in the density-density relation [86].

Consider the density contrasts of visible objects and mass, $\delta_{\text{obj}}(\mathbf{x}, z|R)$ and $\delta_m(\mathbf{x}, z|R)$, at a position \mathbf{x} and a redshift z smoothed over a scale R [87]. In general, the former should depend on various other auxiliary variables $\vec{\mathcal{A}}$ defined at different locations \mathbf{x}' and redshifts z' smoothed over different scales R' in addition to the mass density contrast at the same position, $\delta_m(\mathbf{x}, z|R)$. While this relation can be schematically expressed as

$$\delta_{\text{obj}}(\mathbf{x}, z|R) = \mathcal{F}[\mathbf{x}, z, R, \delta_m(\mathbf{x}, z|R), \vec{\mathcal{A}}(\mathbf{x}', z'|R'), \dots], \quad (118)$$

it is impossible even to specify the list of the astrophysical variables $\vec{\mathcal{A}}$, and thus hopeless to predict the functional form in a rigorous manner. Therefore if one simply focuses on the relation between $\delta_{\text{obj}}(\mathbf{x}, z|R)$ and $\delta_m(\mathbf{x}, z|R)$, the relation becomes inevitably *stochastic* and *nonlinear* due to the dependence on unspecified auxiliary variables $\vec{\mathcal{A}}$.

For illustrative purposes, define the *biasing* factor as the ratio of the density contrasts of luminous objects and mass:

$$B_{\text{obj}}(\mathbf{x}, z|R) \equiv \frac{\delta_{\text{obj}}(\mathbf{x}, z|R)}{\delta_m(\mathbf{x}, z|R)} = \frac{\mathcal{F}[\mathbf{x}, z, R, \delta_m(\mathbf{x}, z|R), \vec{\mathcal{A}}(\mathbf{x}', z'|R'), \dots]}{\delta_m(\mathbf{x}, z|R)}. \quad (119)$$

Only in very idealized situations, the above *nonlocal stochastic nonlinear* factor in terms of δ_m may be approximated by

1. a *local stochastic nonlinear* bias:

$$B_{\text{obj}}(\mathbf{x}, z|R) = b_{\text{obj}}^{(\text{sn})}[\mathbf{x}, z, R, \delta_m(\mathbf{x}, z|R), \vec{\mathcal{A}}(\mathbf{x}, z|R), \dots], \quad (120)$$

2. a *local deterministic nonlinear* bias:

$$B_{\text{obj}}(\mathbf{x}, z|R) = b_{\text{obj}}^{(\text{dn})}[z, R, \delta_m(\mathbf{x}, z|R)], \quad (121)$$

and

3. a *local deterministic linear* bias:

$$B_{\text{obj}}(\mathbf{x}, z|R) = b_{\text{obj}}(z, R) \quad (122)$$

From the above point of view, the local deterministic linear bias is obviously unrealistic, but is still a widely used conventional model for biasing. In fact, the

time- and scale-dependence of the linear bias factor $b_{\text{obj}}(z, R)$ was neglected in many previous studies of biased galaxy formation until very recently. Currently, however, various models beyond the deterministic linear biasing have been seriously considered with particular emphasis on the nonlinear and stochastic aspects of the biasing [71, 15, 86, 87].

4.3 Density peaks and dark matter halos as toy models for galaxy biasing

Let us illustrate the biasing from numerical simulations by considering two specific and popular models; primordial density peaks and dark matter halos [87]. We use the N-body simulation data of $L = 100h^{-1}\text{Mpc}$ again for this purpose [35]. We select density peaks with the threshold of the peak height $\nu_{\text{th}} = 1.0, 2.0$, and 3.0 . As for the dark matter halos, these are identified using the standard friend-of-friend algorithm with a linking length of 0.2 in units of the mean particle separation. We select halos of mass larger than the threshold $M_{\text{th}} = 2.0 \times 10^{12}, 5.0 \times 10^{12}$ and $1.0 \times 10^{13} h^{-1} M_{\odot}$.

Figures 5 and 6 depicts the distribution of dark matter particles (*upper-panel*), peaks (*middle-panel*) and halos (*lower-panel*) in LCDM model at $z = 0$ and $z = 2.2$ within a circular slice (*comoving* radius of $150h^{-1}\text{Mpc}$ and thickness of $15h^{-1}\text{Mpc}$). We locate a fiducial observer in the center of the circle. Then the *comoving* position vector \mathbf{r} for a particle with a *comoving* peculiar velocity \mathbf{v} at a redshift z is observed at the position \mathbf{s} in redshift space:

$$\mathbf{s} = \mathbf{r} + \frac{1}{H(z)} \frac{\mathbf{r} \cdot \mathbf{v}}{|\mathbf{r}|} \frac{\mathbf{r}}{|\mathbf{r}|}, \quad (123)$$

where $H(z)$ is the Hubble parameter at z . The right panels in Figures 5 and 6 plot the observed distribution in redshift space, where the redshift-space distortion is quite visible; the coherent velocity field enhances the structure perpendicular to the line-of-sight of the observer (*squashing*) while the virialized clump becomes elongated along the line-of-sight (*finger-of-God*).

We use two-point correlation functions to quantify stochasticity and nonlinearity in biasing of peaks and halos, and explore the signature of the redshift-space distortion. Since we are interested in the relation of the biased objects and the dark matter, we introduce three different correlation functions; the auto-correlation functions of dark matter and the objects, ξ_{mm} and ξ_{oo} , and their cross-correlation function ξ_{om} . In the present case, the subscript o refers to either h (halos) or ν (peaks). We also use the superscripts R and S to distinguish quantities defined in real and redshift spaces, respectively. We estimate those correlation functions using the standard pair-count method. The correlation function $\xi^{(S)}$ is evaluated under the distant-observer approximation.

Those correlation functions are plotted in Figures 7 and 8 for peaks and halos, respectively. The correlation functions of biased objects generally have

larger amplitudes than those of mass. In nonlinear regimes ($\xi > 1$) the finger-of-god effect suppresses the amplitude of $\xi^{(S)}$ relative to $\xi^{(R)}$, while $\xi^{(S)}$ is larger than $\xi^{(R)}$ in linear regimes ($\xi < 1$) due to the coherent velocity field.

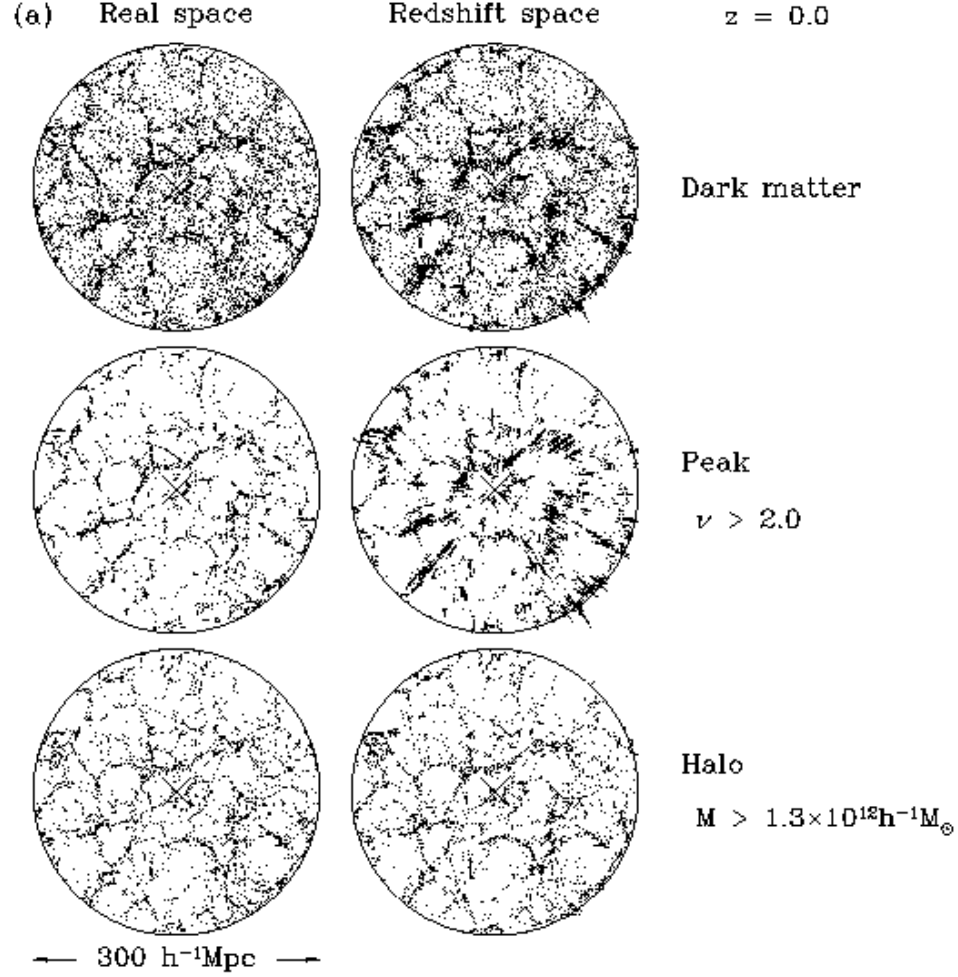


Figure 5: Top-view of distribution of objects at $z = 0$ in real (left panels) and redshift (right panels) spaces around the fiducial observer at the center; dark matter particles (top panels), peaks with $\nu > 2$ (middle panels) and halos with $M > 1.3 \times 10^{12} M_{\odot}$ (bottom panels) in LCDM model. The thickness of those slices is $15 h^{-1} \text{Mpc}$. (ref. [87])

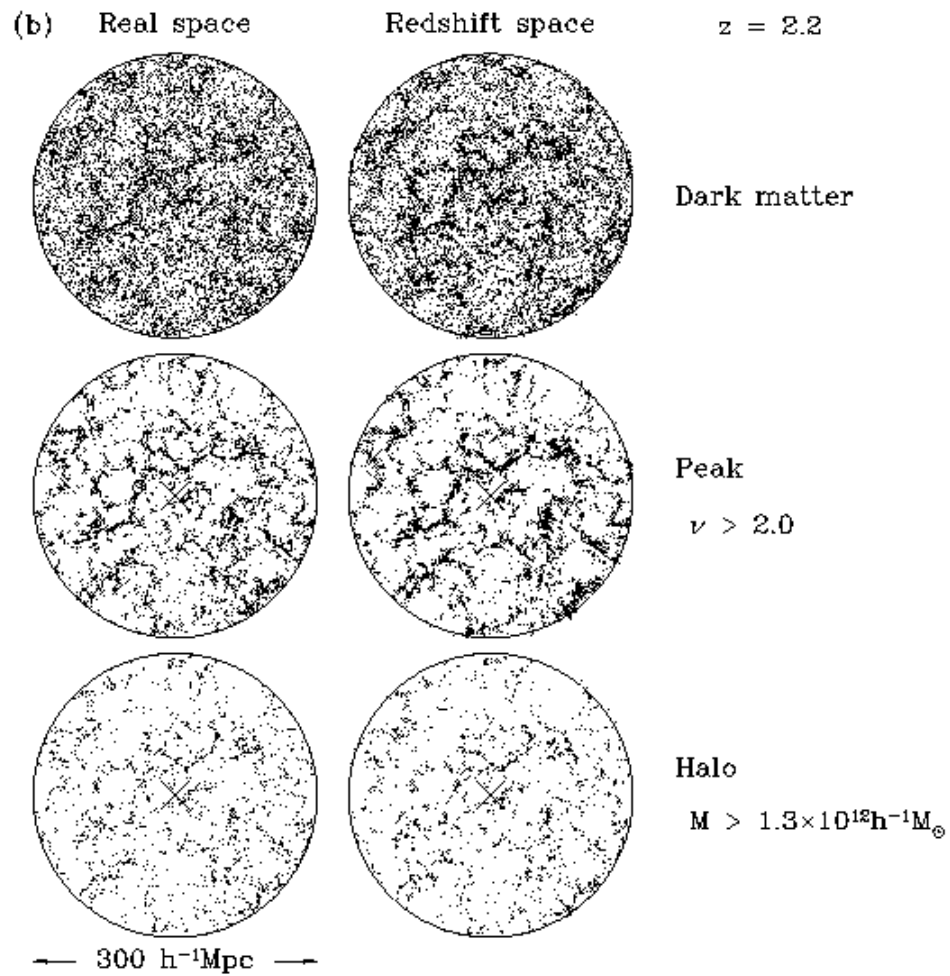


Figure 6: Same as Fig.5 but at $z = 2.2$. (ref. [87])

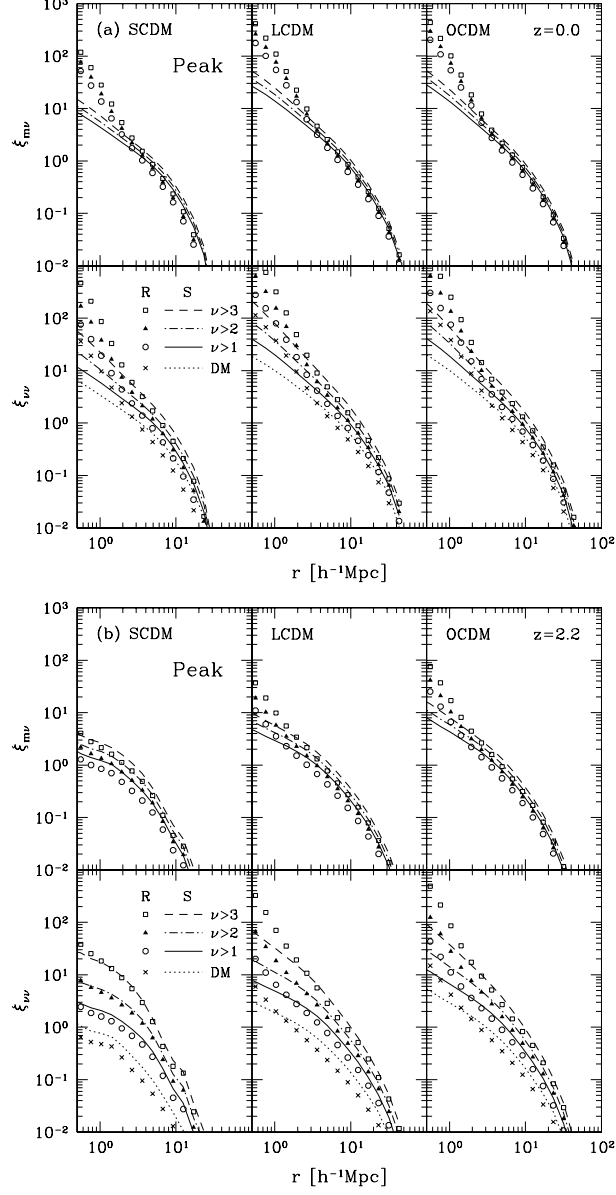


Figure 7: Auto- and cross-correlation functions of dark matter and peaks in SCDM (left panels), LCDM (middle panels) and OCDM (right panels). Different symbols indicate the results in real space (open squares for $\nu > 3$, filled triangles for $\nu > 2$, open circles for $\nu > 1$, and crosses for dark matter), while different curves indicate those in redshift space (dashed for $\nu > 3$, dot-dashed for $\nu > 2$, solid for $\nu > 1$, and dotted for dark matter). (a) $z = 0$, (b) $z = 2.2$. (ref. [87])

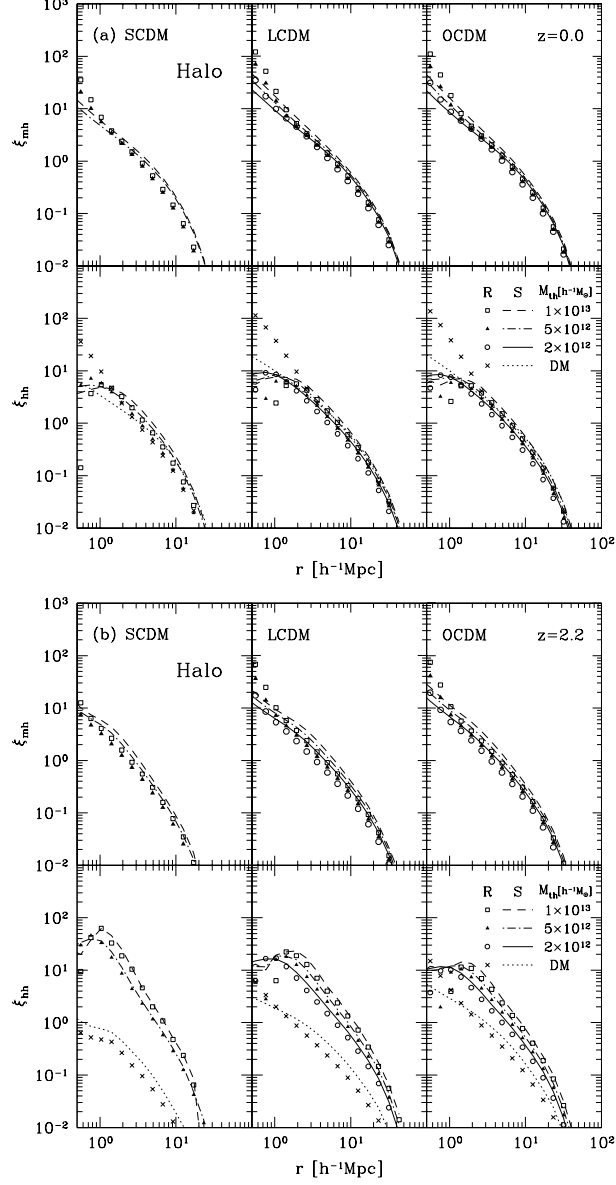


Figure 8: Same as Fig. 7 but for halo model; open squares and dashed lines for $M > 10^{13} h^{-1} M_{\odot}$, filled triangles and dot-dashed lines for $M > 5 \times 10^{12} h^{-1} M_{\odot}$, open circles and solid lines for $M > 2 \times 10^{12} h^{-1} M_{\odot}$, and crosses and dotted lines for dark matter. For SCDM model, we only plot the correlation functions with $M_{\text{th}} = 5 \times 10^{12}, 10^{13} h^{-1} M_{\odot}$. (a) $z = 0$, (b) $z = 2.2$. (ref. [87])

4.4 Biasing of galaxies in cosmological hydrodynamic simulations

Popular models of the biasing based on the peak or the dark halos are successful in capturing some essential features of biasing. None of the existing models of bias, however, seems to be sophisticated enough for the coming precision cosmology era. Development of a more detailed theoretical model of bias is needed. A straightforward next step is to resort to numerical simulations which take account of galaxy formation even if phenomenological at this point. We show an example of such approaches from Yoshikawa et al. (2001) [91] who apply cosmological smoothed particle hydrodynamic (SPH) simulations in LCDM model with particular attention to the comparison of the biasing of dark halos and simulated galaxies (see also ref. [78]).

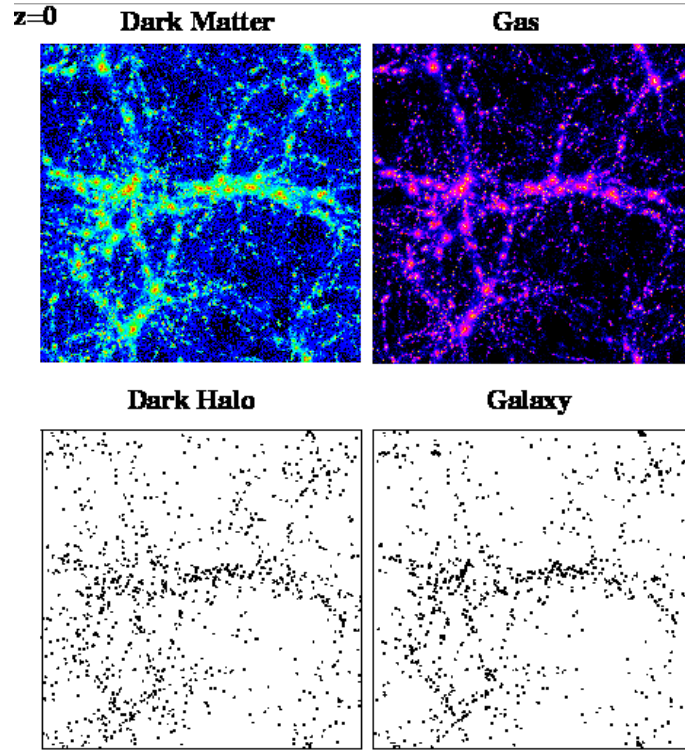


Figure 9: Distribution of gas particles, dark matter particles, galaxies and dark halos in the volume of $75h^{-1} \times 75h^{-1} \times 30h^{-1} \text{Mpc}^3$ model at $z = 0$. *Upper-right*: gas particles; *Upper-left*: dark matter particles; *Lower-right*: galaxies; *Lower-left*: DM cores. (ref. [91])

Galaxies in their simulations are identified as clumps of cold and dense gas particles which satisfy the Jeans condition and have the SPH density more than 100 times the mean baryon density at each redshift. Dark halos are identified with a standard friend-of-friend algorithm; the linking length is 0.164 times the mean separation of dark matter particles, for instance, at $z = 0$. In addition, they identify the surviving high-density substructures in dark halos, DM cores. See ref. [91] for further details.

Figure 9 illustrates the distribution of dark matter particles, gas particles, dark halos and galaxies at $z = 0$ where galaxies are more strongly clustered than dark halos. Figure 10 depicts a close-up snapshot of the most massive cluster at $z = 0$ with mass $M \simeq 8 \times 10^{14} M_{\odot}$. Circles in lower panels indicate the positions of galaxies identified in our simulation.

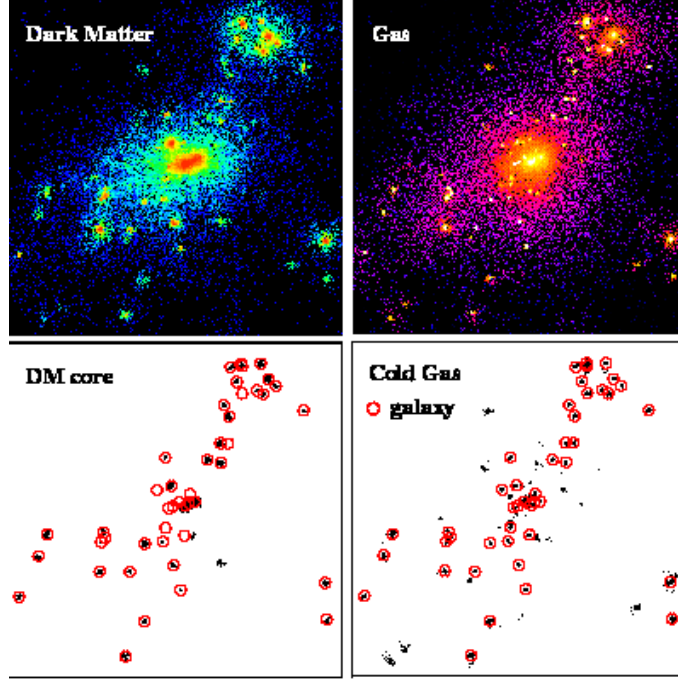


Figure 10: Snapshots of the most massive cluster ($M \simeq 8 \times 10^{14} M_{\odot}$) in the simulation at $z = 0$. *Upper-left*: dark matter; *Upper-right*: gas; *Lower-left*: DM cores; *Lower-right*: cold gas. Circles in lower panels indicate the positions of galaxies identified according to our criteria. The comoving size of the box is $6.25h^{-1}\text{Mpc}$ per side. (ref. [91])

Figure 11 shows the joint distribution of δ_h and δ_g with mass density field

δ_m at redshift $z = 0, 1$ and 2 smoothed over $R_s = 12h^{-1}\text{Mpc}$. The conditional mean relation $\bar{\delta}_i(\delta_m)$ computed directly from the simulation is plotted in solid lines while dashed lines indicate theoretical predictions of halo biasing by Taruya & Suto (2000) [86]. For a given smoothing scale, the simulated halos exhibit positive biasing for relatively small δ_m in agreement with the predictions. On the other hand, they tend to be underpopulated for large δ_m , or *anti-biased*. This is mainly due to the exclusion effect of dark halos due to their finite volume size which is not taken into account in the theoretical model. Since our simulated *galaxies* have smaller spatial extent than the halos, the exclusion effect is not so serious. This is clearly illustrated in lower panels in Figure 11, and indeed they show much better agreement with the theoretical model.

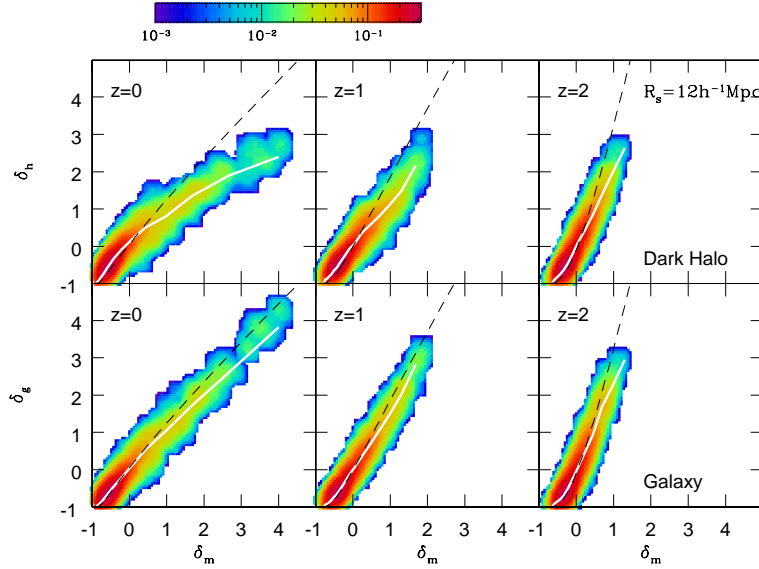


Figure 11: Joint probability distributions of overdensity fields for dark halos and galaxies with dark matter overdensity smoothed over $R_s = 12h^{-1}\text{Mpc}$ at redshift $z = 0, 1$ and 2 . Solid lines indicate the conditional mean $\bar{\delta}_i(\delta_m)$ for each object. Dashed lines in each panel depict the theoretical prediction of conditional mean by Taruya & Suto (2000). (ref. [91])

Turn next to a more conventional biasing parameter defined through the two-point statistics:

$$b_{\xi,i}(r) \equiv \sqrt{\frac{\xi_{ii}(r)}{\xi_{mm}(r)}}, \quad (124)$$

where $\xi_{ii}(r)$ and $\xi_{mm}(r)$ are two-point correlation functions of objects i and

of dark matter, respectively. While the above biasing parameter is ill-defined where either $\xi_{ii}(r)$ or $\xi_{mm}(r)$ becomes negative, it is not the case at clustering scales of interest ($< 10h^{-1}\text{Mpc}$).

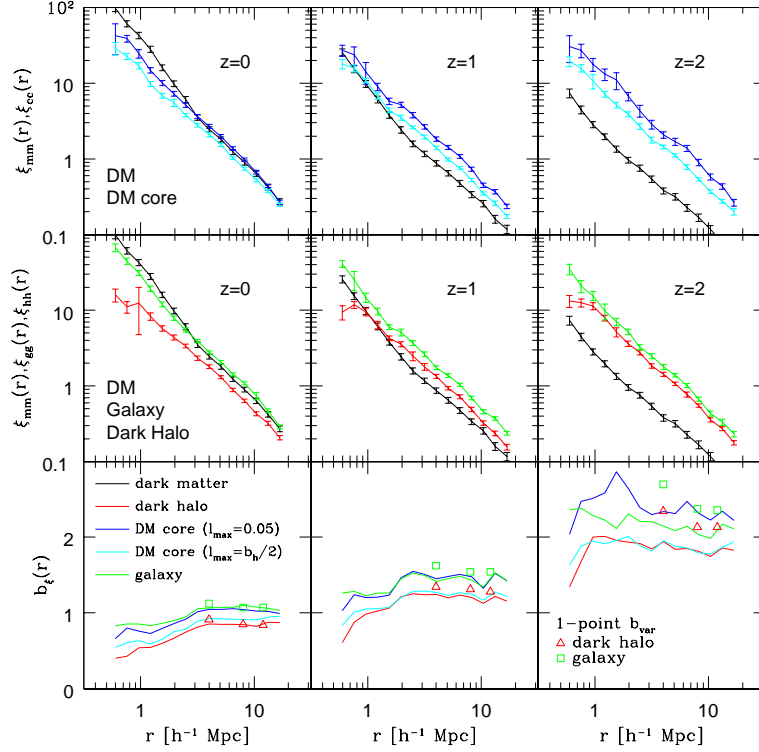


Figure 12: Two-point correlation functions of dark matter, galaxies, and dark halos from cosmological hydrodynamical simulations. (ref. [91])

Figure 12 shows two-point correlation functions of dark matter, galaxies, dark halos and DM cores (*upper and middle panels*), and the profiles of biasing parameters $b_\xi(r)$ for those objects (*lower panels*) at $z = 0, 1$ and 2 . In lower panels, we also plot the parameter $b_{\text{var},i} \equiv \sigma_i/\sigma_m$, which are defined in terms of the one-point statistics (variance), for comparison on the smoothing scale $R_s = 4h^{-1}\text{Mpc}$, $8h^{-1}\text{Mpc}$ and $12h^{-1}\text{Mpc}$ at $r = R_s$ for each kind of objects by different symbols. In the upper panels, we show the correlation functions of DM cores identified with two different maximum linking length; $l_{\text{max}} = 0.05$ and $l_{\text{max}} = b_h/2$. Correlation functions of DM cores identified with $l_{\text{max}} = 0.05$ are similar to those of galaxies. On the other hand, those identified with $l_{\text{max}} = b_h/2$ exhibit much weaker correlation, and are rather similar to those of dark halos.

This is due to the fact that the present algorithm of group identification with larger l_{max} tends to pick up lower mass halos which are poorly resolved in our numerical resolution.

The correlation functions of galaxies are almost unchanged with redshift, and correlation functions of dark halos only slightly evolve between $z = 0$ and 2. By contrast, the amplitude of the dark matter correlation functions evolves rapidly by factor of ~ 10 from $z = 2$ to $z = 0$. The biasing parameter $b_{\xi, \text{g}}$ is larger at a higher redshift, for example, $b_{\xi, \text{g}} \simeq 2\text{--}2.5$ at $z = 2$. The biasing parameter $b_{\xi, \text{h}}$ for dark halos is systematically lower than that of galaxies and DM cores again due to the volume exclusion effect. At $z = 0$, galaxies and DM cores are slightly anti-biased relative to dark matter at $r \simeq 1 h^{-1} \text{Mpc}$. In lower panels, we also plot the one-point biasing parameter $b_{\text{var}, i} \equiv \sigma_i / \sigma_{\text{m}}$ at $r = R_s$ for comparison. In general we find that $b_{\xi, i}$ is very close to $b_{\text{var}, i}$ at $z \sim 0$, but systematically lower than $b_{\text{var}, i}$ at higher redshifts.

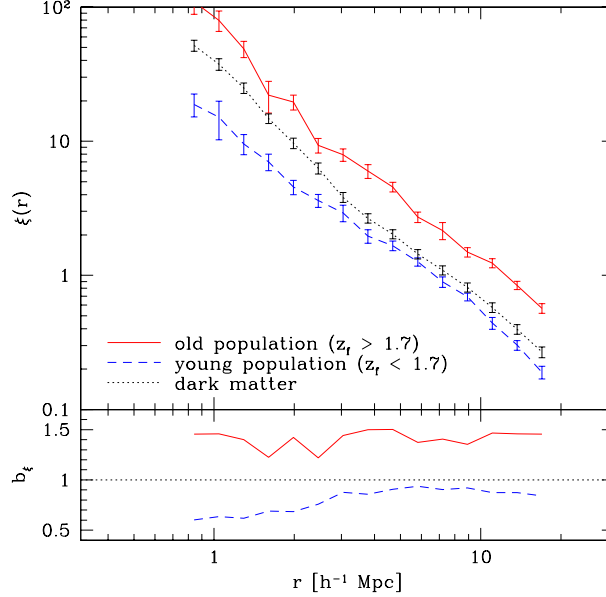


Figure 13: Two-point correlation functions for the old and young populations of galaxies at $z = 0$ as well as that of dark matter distribution. The profiles of bias parameters $b_{\xi}(r)$ for both of the two populations are also shown in the lower panel. (ref. [91])

For each galaxy identified at $z = 0$, we define its formation redshift z_f by the epoch when half of its *cooled gas* particles satisfy our criteria of galaxy formation. Roughly speaking, z_f corresponds to the median formation redshift of *stars* in

the present-day galaxies. We divide all simulated galaxies at $z = 0$ into two populations (the young population with $z_f < 1.7$ and the old population with $z_f > 1.7$) so as to approximate the observed number ratio of 3/1 for late-type and early-type galaxies.

The difference of the clustering amplitude can be also quantified by their two-point correlation functions at $z = 0$ as plotted in Figure 13. The old population indeed clusters more strongly than the mass, and the young population is anti-biased. The relative bias between the two populations $b_{\xi,g}^{\text{el}} \equiv \sqrt{\xi_{\text{old}}/\xi_{\text{young}}}$ ranges 1.5 and 2 for $1h^{-1}\text{Mpc} < r < 20h^{-1}\text{Mpc}$, where ξ_{young} and ξ_{old} are the two-point correlation functions of the young and old populations.

4.5 Halo occupation function approach for galaxy biasing

Since the clustering of dark matter halos is well understood now, one can describe the galaxy biasing if the halo model is combined with the relation between the halos and luminous objects. This is another approach to galaxy biasing, *halo occupation function* (HOF), which becomes very popular recently. Indeed the basic idea behind HOF has a long history, but the model predictions have been significantly improved with the recent accurate models for the mass function, the biasing and the density profile of dark matter halos. We refer the readers for an extensive review on the HOF by Cooray & Sheth (2002) [12]. Here we briefly outline this approach.

We adopt a simple parametric form for the average number of a given galaxy population as a function of the hosting halo mass:

$$N_g(M) = \begin{cases} (M/M_1)^\alpha & (M > M_{\text{min}}) \\ 0 & (M < M_{\text{min}}) \end{cases}. \quad (125)$$

The above statistical and empirical relation is the essential ingredient in the current modeling characterized by the minimum mass of halos which host the population of galaxies (M_{min}), a normalization parameter which can be interpreted as the critical mass above which halos typically host more than one galaxy (M_1 ; note that M_1 may exceed M_{min} since the above relation represents the statistical expected value of number of galaxies), and the power-law index of the mass dependence of the efficiency of galaxy formation (α). We will put constraints on the three parameters from the observed number density and clustering amplitude for each galaxy population. In short, the number density of galaxies is most sensitive to M_1 which changes the average number of galaxies per halo. The clustering amplitude on large scales is determined by the hosting halos and thus very sensitive to the mass of those halos, M_{min} . The clustering on smaller scales, on the other hand, depends on those three parameters in a fairly complicated fashion; roughly speaking, M_{min} changes the amplitude, while α , and to a lesser extent M_1 as well, changes the slope.

With the above relation, the number density of the corresponding galaxy

population at redshift z is given by

$$n_{g,z}(z) = \int_{M_{\min}}^{\infty} dM n_{\text{halo}}(M, z) N_g(M), \quad (126)$$

where $n_{\text{halo}}(M)$ denotes the halo mass function.

The galaxy two-point correlation function on small scales is dominated by contributions of galaxy pairs located in the same halo. For instance, Bullock et al. (2002) [8] adopted the mean number of galaxy pairs $\langle N_g(N_g - 1) \rangle(M)$ within a halo of mass M of the form:

$$\langle N_g(N_g - 1) \rangle(M) = \begin{cases} N_g^2(M) & (N_g(M) > 1) \\ N_g^2(M) \log(4N_g(M)) / \log(4) & (1 > N_g(M) > 0.25) \\ 0 & (N_g(M) < 0.25) \end{cases}. \quad (127)$$

In the framework of the halo model, the galaxy power spectrum consists of two contributions, one from galaxy pairs located in the same halo (1-halo term) and the other from galaxy pairs located in two different halos (2-halo term):

$$P_g^{\text{tot}}(k) = P_g^{1h}(k) + P_g^{2h}(k). \quad (128)$$

The 1-halo term is written as

$$P_g^{1h}(k) = \frac{1}{(2\pi)^3 n_{g,z}^2} \int dM n_{\text{halo}}(M) \langle N_g(N_g - 1) \rangle b(M) |y(k, M)|^p. \quad (129)$$

Seljak (2000) [77] chose $p = 2$ for $\langle N_g(N_g - 1) \rangle > 1$ and $p = 1$ for $\langle N_g(N_g - 1) \rangle < 1$. The 2-halo term on the assumption of the linear halo bias model [59] reduces to

$$P_g^{2h}(k) = \frac{P_{\text{lin}}(k)}{n_{g,z}^2} \left[\int dM n_{\text{halo}}(M) N_g(M) b(M) y(k, M) \right]^2, \quad (130)$$

where $P_{\text{lin}}(k)$ is the linear dark matter power spectrum, $b(M)$ is the halo bias factor, and $y(k, M)$ is the Fourier transform of the halo dark matter profile normalized by its mass, $y(k, M) = \tilde{\rho}(k, M)/M$ [77].

The halo occupation formalism, although simple, provides a useful framework in deriving constraints on galaxy formation models from large data sets of the upcoming galaxy redshift surveys. For example, Zehavi et al. (2003) [99] used the halo occupation formalism to model departures from a power law in the SDSS galaxy correlation function. They demonstrated that this is due to the transition from a large-scale regime dominated by galaxy pairs in different halos to a small-scale regime dominated by those in the same halo. Magliocchetti & Porciani (2003) [47] applied the halo occupation formalism to the 2dFGRS clustering results per spectral type of Madgwick et al. (2003) [45]. This provides constraints on the distribution of late-type and early-type galaxies within the dark matter halos of different mass.

5 Relativistic effects observable in clustering at high redshifts

Redshift surveys of galaxies definitely serve as the central database for observational cosmology. In addition to the existing *shallower* surveys ($z < 0.2$), clustering in the universe in the range $z = 1 - 3$ has been partially revealed by, for instance, the Lyman-break galaxies and X-ray selected AGNs. In particular, the 2dF and SDSS QSO redshift surveys promise to extend the observable scale of the universe by an order of magnitude, up to a few Gpc. A proper interpretation of such redshift surveys in terms of the clustering evolution, however, requires an understanding of many cosmological effects which can be neglected for $z \ll 1$ and thus have not been considered seriously so far. These cosmological *contaminations* include linear redshift-space (velocity) distortion, nonlinear redshift-space (velocity) distortion, cosmological redshift-space (geometrical) distortion, and cosmological light-cone effect.

We describe a theoretical formalism to incorporate those effects, in particular the cosmological redshift-distortion and light-cone effects, and present several specific predictions in CDM models. The details of the material presented in this section may be found in refs. [83, 89, 90, 46, 28, 29].

5.1 Cosmological lightcone effect on the two-point correlation functions

Observing a distant patch of the universe is equivalent to observing the past. Due to the finite light velocity, a line-of-sight direction of a redshift survey is along the time, as well as spatial, coordinate axis. Therefore the entire sample does not consist of objects on a constant-time hypersurface, but rather on a light-cone, i.e., a null hypersurface defined by observers at $z = 0$. This implies that many properties of the objects change across the depth of the survey volume, including the mean density, the amplitude of spatial clustering of dark matter, the bias of luminous objects with respect to mass, and the intrinsic evolution of the absolute magnitude and spectral energy distribution. These aspects should be properly taken into account in order to extract cosmological information from observed samples of redshift surveys.

In order to predict quantitatively the two-point statistics of objects on the light cone, one must take account of (i) nonlinear gravitational evolution, (ii) linear redshift-space distortion, (iii) nonlinear redshift-space distortion, (iv) weighted averaging over the light-cone, (v) cosmological redshift-space distortion due to the geometry of the universe, and (vi) object-dependent clustering bias. The effect (v) comes from our ignorance of the correct cosmological parameters, and (vi) is rather sensitive to the objects which one has in mind. Thus the latter two effects will be discussed in the next subsection.

Nonlinear gravitational evolution of mass density fluctuations is now well

understood, at least for two-point statistics. In practice, we adopt an accurate fitting formula [65] for the nonlinear power spectrum $P_{\text{nl}}^{\text{R}}(k, z)$ in terms of its linear counterpart. If one assumes a scale-independent deterministic linear bias, furthermore, the power spectrum distorted by the peculiar velocity field is known to be well approximated by the following expression:

$$P^{(\text{S})}(k_{\perp}, k_{\parallel}; z) = b^2(z) P_{\text{mass}}^{(\text{R})}(k; z) \left[1 + \beta(z) \left(\frac{k_{\parallel}}{k} \right)^2 \right]^2 D_{\text{vel}}[k_{\parallel} \sigma_{\text{P}}(z)], \quad (131)$$

where k_{\perp} and k_{\parallel} are the comoving wavenumber perpendicular and parallel to the line-of-sight of an observer, and $P_{\text{mass}}^{(\text{R})}(k; z)$ is the mass power spectrum in real space. The second factor in the right-hand-side comes from the linear redshift-space distortion [38], and the last factor is a phenomenological correction for non-linear velocity effect [65]. In the above, we introduce

$$\beta(z) \equiv \frac{1}{b(z)} \frac{d \ln D(z)}{d \ln a} \simeq \frac{1}{b(z)} \left\{ \Omega^{0.6}(z) + \frac{\lambda(z)}{70} \left[1 + \frac{\Omega(z)}{2} \right] \right\}. \quad (132)$$

We assume that the pair-wise velocity distribution in real space is approximated by

$$f_v(v_{12}) = \frac{1}{\sqrt{2}\sigma_{\text{P}}} \exp \left(-\frac{\sqrt{2}|v_{12}|}{\sigma_{\text{P}}} \right), \quad (133)$$

with σ_{P} being the 1-dimensional pair-wise peculiar velocity dispersion. Then the finger-of-god effect is modeled by the damping function, $D_{\text{vel}}[k_{\parallel} \sigma_{\text{P}}(z)]$:

$$D_{\text{vel}}[k \mu \sigma_{\text{P}}] = \frac{1}{1 + \kappa^2 \mu^2}, \quad (134)$$

where μ is the direction cosine in k -space, and the dimensionless wavenumber κ is related to the peculiar velocity dispersion σ_{P} in the physical velocity units:

$$\kappa(z) = \frac{k(1+z)\sigma_{\text{P}}(z)}{\sqrt{2}H(z)}. \quad (135)$$

Since we are mainly interested in the scales around $1h^{-1}\text{Mpc}$, we adopt the following fitting formula throughout the analysis below which better approximates the small-scale dispersions in physical units:

$$\sigma_{\text{P}}(z) \sim \begin{cases} 740(1+z)^{-1} \text{km/s} & \text{for } \Lambda\text{CDM model} \\ 650(1+z)^{-0.8} \text{km/s} & \text{for } \Lambda\text{CDM model.} \end{cases} \quad (136)$$

Integrating equation (131) over μ , one obtains the direction-averaged power spectrum in redshift space:

$$\frac{P_{\text{nl}}^{\text{S}}(k, z)}{P_{\text{nl}}^{\text{R}}(k, z)} = A(\kappa) + \frac{2}{3}\beta(z)B(\kappa) + \frac{1}{5}\beta^2(z)C(\kappa) \quad (137)$$

where

$$A(\kappa) = \frac{\arctan(\kappa)}{\kappa}, \quad (138)$$

$$B(\kappa) = \frac{3}{\kappa^2} \left[1 - \frac{\arctan(\kappa)}{\kappa} \right], \quad (139)$$

$$C(\kappa) = \frac{5}{3\kappa^2} \left[1 - \frac{3}{\kappa^2} + \frac{3 \arctan(\kappa)}{\kappa^3} \right]. \quad (140)$$

Adopting those approximations, the direction-averaged correlation functions on the light-cone are finally computed as

$$\xi^{\text{LC}}(x_s) = \frac{\int_{z_{\min}}^{z_{\max}} dz \frac{dV_c}{dz} [\phi(z)n_0(z)]^2 \xi(x_s; z)}{\int_{z_{\min}}^{z_{\max}} dz \frac{dV_c}{dz} [\phi(z)n_0(z)]^2}, \quad (141)$$

where z_{\min} and z_{\max} denote the redshift range of the survey, and

$$\xi(x_s; z) \equiv \frac{1}{2\pi^2} \int_0^\infty P_{\text{nl}}^{\text{S}}(k, z) \frac{\sin kx_s}{kx_s} k^2 dk. \quad (142)$$

Throughout the present analysis, we assume a standard Robertson – Walker metric of the form:

$$ds^2 = -dt^2 + a(t)^2 \{d\chi^2 + S_K(\chi)^2 [d\theta^2 + \sin^2 \theta d\phi^2]\}, \quad (143)$$

where $S_K(\chi)$ is determined by the sign of the curvature K as

$$S_K(\chi) = \begin{cases} \sin(\sqrt{K}\chi)/\sqrt{K} & (K > 0) \\ \chi & (K = 0) \\ \sinh(\sqrt{-K}\chi)/\sqrt{-K} & (K < 0) \end{cases}, \quad (144)$$

where the present scale factor a_0 is normalized as unity, and the spatial curvature K is given as (eq.[13])

$$K = H_0^2(\Omega_m + \Omega_\Lambda - 1). \quad (145)$$

The radial comoving distance $\chi(z)$ is computed by

$$\chi(z) = \int_t^{t_0} \frac{dt}{a(t)} = \int_0^z \frac{dz}{H(z)}. \quad (146)$$

The comoving angular diameter distance $D_c(z)$ at redshift z is equivalent to $S^{-1}(\chi(z))$, and, in the case of $\Omega_\Lambda = 0$, is explicitly given by Mattig's formula:

$$D_c(z) = \frac{1}{H_0} \frac{z}{1+z} \frac{1+z+\sqrt{1+\Omega_m z}}{1+\Omega_m z/2+\sqrt{1+\Omega_m z}}. \quad (147)$$

Then dV_c/dz , the comoving volume element per unit solid angle, is explicitly given as

$$\begin{aligned}\frac{dV_c}{dz} &= S_K^2(\chi) \frac{d\chi}{dz} \\ &= \frac{S_K^2(\chi)}{H_0 \sqrt{\Omega_m(1+z)^3 + (1 - \Omega_m - \Omega_\Lambda)(1+z)^2 + \Omega_\Lambda}}.\end{aligned}\quad (148)$$

5.2 Evaluating two-point correlation functions from N-body simulation data

The theoretical modeling described above was tested against simulation results by Hamana, Colombi & Suto (2001) [28]. Using cosmological N -body simulations in SCDM and Λ CDM models, they generated light-cone samples as follows; first, they adopt a distance observer approximation and assume that the line-of-sight direction is parallel to Z -axis regardless with its (X, Y) position. Second, they periodically duplicate the simulation box along the Z -direction so that at a redshift z , the position and velocity of those particles locating within an interval $\chi(z) \pm \Delta\chi(z)$ are dumped, where $\Delta\chi(z)$ is determined by the output time-interval of the original N -body simulation. Finally they extract five independent (non-overlapping) cone-shape samples with the angular radius of 1 degree (the field-of-view of π degree²). In this manner, they have generated mock data samples on the light-cone continuously extending up to $z = 0.4$ (relevant for galaxy samples) and $z = 2.0$ (relevant for QSO samples), respectively from the small and large boxes.

Two-point correlation function is estimated by the conventional pair-count adopting the following estimator [43]:

$$\xi(x) = \frac{DD(x) - 2DR(x) + RR(x)}{RR(x)}.\quad (149)$$

The comoving separation x_{12} of two objects located at z_1 and z_2 with an angular separation θ_{12} is given by

$$\begin{aligned}x_{12}^2 &= x_1^2 + x_2^2 - Kx_1^2x_2^2(1 + \cos^2\theta_{12}) \\ &\quad - 2x_1x_2\sqrt{1 - Kx_1^2}\sqrt{1 - Kx_2^2}\cos\theta_{12},\end{aligned}\quad (150)$$

where $x_1 \equiv D_c(z_1)$ and $x_2 \equiv D_c(z_2)$.

In redshift space, the observed redshift z_{obs} for each object differs from the “real” one z_{real} due to the velocity distortion effect:

$$z_{\text{obs}} = z_{\text{real}} + (1 + z_{\text{real}})v_{\text{pec}},\quad (151)$$

where v_{pec} is the line of sight relative peculiar velocity between the object and the observer in *physical* units. Then the comoving separation s_{12} of two objects

in redshift space is computed as

$$s_{12}^2 = s_1^2 + s_2^2 - K s_1^2 s_2^2 (1 + \cos^2 \theta_{12}) - 2s_1 s_2 \sqrt{1 - K s_1^2} \sqrt{1 - K s_2^2} \cos \theta_{12}, \quad (152)$$

where $s_1 \equiv D_c(z_{\text{obs},1})$ and $s_2 \equiv D_c(z_{\text{obs},2})$.

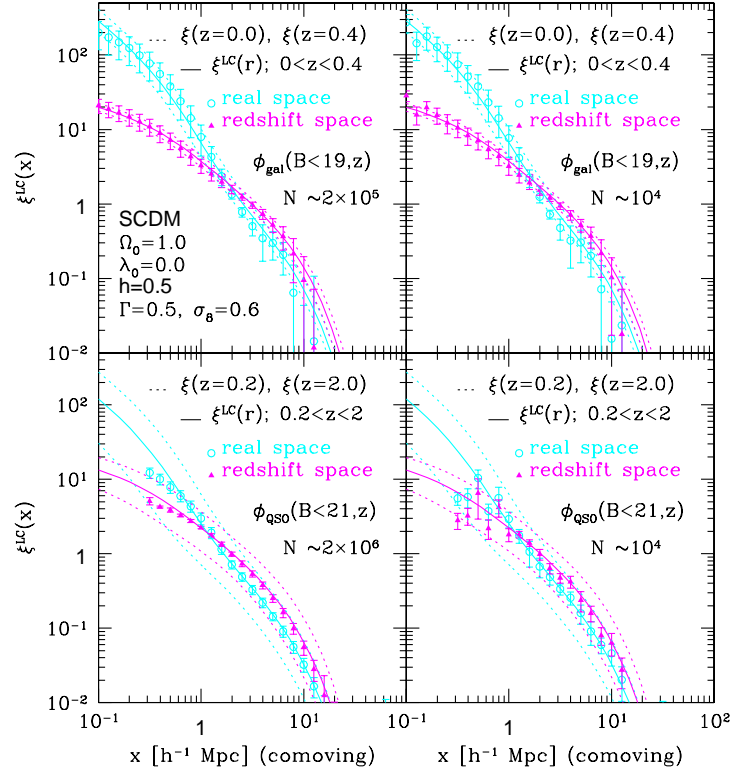


Figure 14: Mass two-point correlation functions on the light cone for particles with redshift-dependent selection functions in SCDM model. *Upper*: $z < 0.4$, *Lower*: $0.2 < z < 2.0$. *Left*: with selection function whose shape is the same as that of the B-band magnitude limit of 19 for galaxies (upper) and 21 for QSOs (lower). *Right*: randomly selected $N \sim 10^4$ particles from the particles in the left results. (ref. [28])

In properly predicting the power spectra on the light cone, the selection function should be specified. For galaxies, they adopt a B-band luminosity function of the APM galaxies fitted to the Schechter function [44]. For quasars, they

adopt the B-band luminosity function from the 2dF QSO survey data [7]. To compute the B-band apparent magnitude from a quasar of absolute magnitude M_B at z (with the luminosity distance $d_L(z)$), they applied the K-correction:

$$B = M_B + 5 \log(d_L(z)/10\text{pc}) - 2.5(1 - p) \log(1 + z) \quad (153)$$

for the quasar energy spectrum $L_\nu \propto \nu^{-p}$ (they used $p = 0.5$). In practice, they adopt the galaxy selection function $\phi_{\text{gal}}(< B_{\text{lim}}, z)$ with $B_{\text{lim}} = 19$ and $z_{\text{min}} = 0.01$ for the small box realizations, and the QSO selection function $\phi_{\text{QSO}}(< B_{\text{lim}}, z)$ with $B_{\text{lim}} = 21$ and $z_{\text{min}} = 0.2$ for the large box realizations. They do not introduce the spatial biasing between selected particles and the underlying dark matter.

Figures 14 and 15 plot the two-point correlation functions in SCDM and Λ CDM, respectively, taking account of the selection functions. It is clear that the simulation results and the predictions are in good agreement.

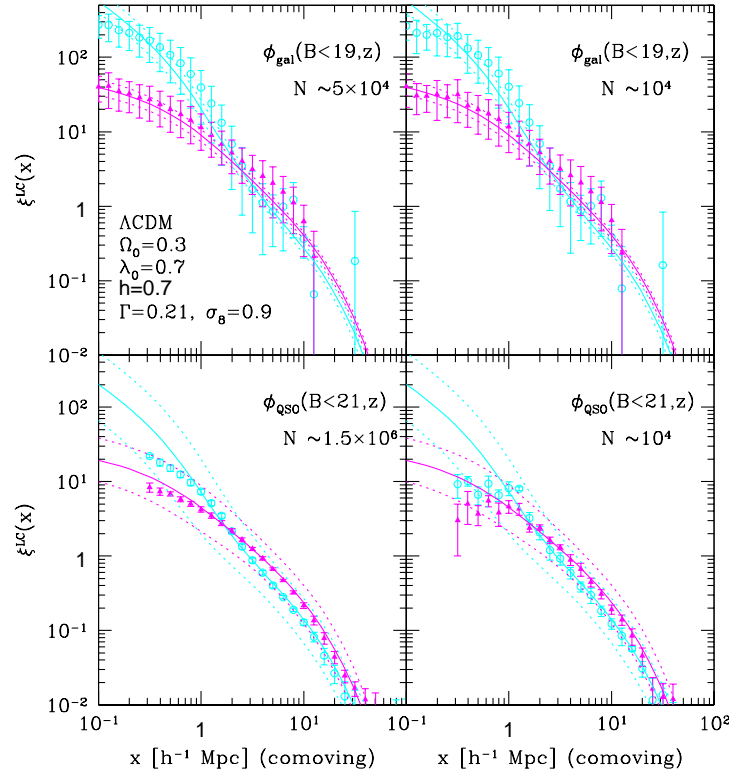


Figure 15: Same as Figure 14 but for Λ CDM model. (ref. [28])

5.3 Cosmological redshift-space distortion

Consider a spherical object at high redshift. If the wrong cosmology is assumed in interpreting the distance-redshift relation along the line of sight and in the transverse direction, the sphere will appear distorted. Alcock & Paczynski (1979) [2] pointed out that this curvature effect could be used to estimate the cosmological constant. Matsubara & Suto (1996) [54] and Ballinger, Peacock & Heavens (1996) [4] developed a theoretical framework to describe the geometrical distortion effect (cosmological redshift distortion) in the two-point correlation function and the power spectrum of distant objects, respectively. Certain studies were less optimistic than others about the possibility of measuring this AP effect. For example, Ballinger, Peacock and Heavens (1996) argued that the geometrical distortion could be confused with the dynamical redshift distortions caused by peculiar velocities and characterized by the linear theory parameter $\beta \equiv \frac{\Omega_m^{0.6}}{b}$. Matsubara & Szalay (2002, 2003) [55, 56] showed that the typical SDSS and 2dF samples of normal galaxies at low redshift (~ 0.1) have sufficiently low signal-to-noise, but they are too shallow to detect the Alcock & Paczynski effect. On the other hand, the quasar SDSS and 2dFGRS surveys are at a useful redshift, but they are too sparse. A more promising sample is the SDSS Luminous Red Galaxies survey (out to redshift $z \sim 0.5$) which turns out to be optimal in terms of both depth and density.

While this analysis is promising, it remains to be tested if non-linear clustering and complicated biasing (which is quite plausible for red galaxies) would not ‘contaminate’ the measurement of the Equation of State. Even if the Alcock & Paczynski test turns out to be less accurate than other cosmological tests (e.g., CMB and SN Ia) the effect itself is an interesting and important ingredient in analyzing the clustering pattern of galaxies at high redshifts. We shall now present the formalism for this effect.

Due to a general-relativistic effect through the geometry of the universe, the *observable* separations perpendicular and parallel to the line-of-sight direction, $x_{s\perp} = (c/H_0)z\delta\theta$ and $x_{s\parallel} = (c/H_0)\delta z$, are mapped differently to the corresponding comoving separations in real space x_\perp and x_\parallel :

$$x_{s\perp}(z) = x_\perp cz / [H_0(1+z)d_A(z)] \equiv x_\perp / c_\perp(z), \quad (154)$$

$$x_{s\parallel}(z) = x_\parallel H(z) / H_0 \equiv x_\parallel / c_\parallel(z), \quad (155)$$

with $d_A(z)$ being the angular diameter distance. The difference between $c_\perp(z)$ and $c_\parallel(z)$ generates an apparent anisotropy in the clustering statistics, which should be isotropic in the comoving space. Then the power spectrum in cosmological redshift space, $P^{(\text{CRD})}$, is related to $P^{(\text{S})}$ defined in the *comoving* redshift space as

$$P^{(\text{CRD})}(k_{s\perp}, k_{s\parallel}; z) = \frac{1}{c_\perp(z)^2 c_\parallel(z)} P^{(\text{S})}\left(\frac{k_{s\perp}}{c_\perp(z)}, \frac{k_{s\parallel}}{c_\parallel(z)}; z\right), \quad (156)$$

where the first factor comes from the Jacobian of the volume element $dk_{s\perp}^2 dk_{s\parallel}$, and $k_{s\perp} = c_{\perp}(z)k_{\perp}$ and $k_{s\parallel} = c_{\parallel}(z)k_{\parallel}$ are the wavenumber perpendicular and parallel to the line-of-sight direction.

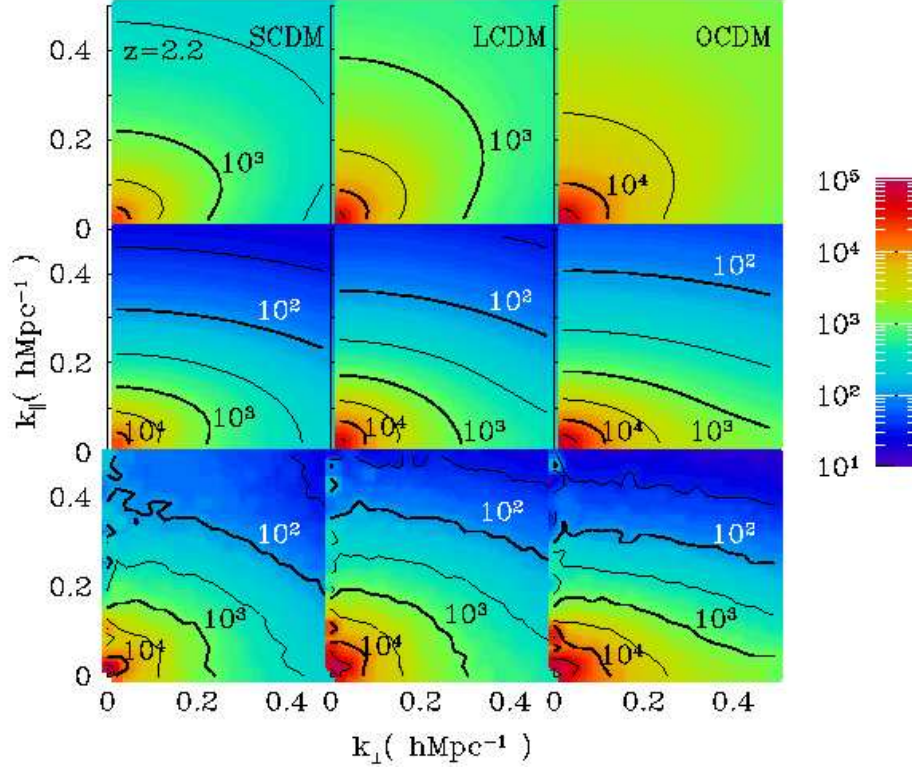


Figure 16: Two-dimensional power spectra in cosmological redshift space at $z = 2.2$. (ref. [46])

Using equation (131), equation (156) reduces to

$$P^{(\text{CRD})}(k_s, \mu_k; z) = \frac{b^2(z)}{c_{\perp}(z)^2 c_{\parallel}(z)} P_{\text{mass}}^{(\text{R})} \left(\frac{k_s}{c_{\perp}(z)} \sqrt{1 + \left[\frac{1}{\eta(z)^2} - 1 \right] \mu_k^2}; z \right) \times \left\{ 1 + \left[\frac{1}{\eta(z)^2} - 1 \right] \mu_k^2 \right\}^{-2} \left\{ 1 + \left[\frac{1 + \beta(z)}{\eta(z)^2} - 1 \right] \mu_k^2 \right\}^2 \left[1 + \frac{k_s^2 \mu_k^2 \sigma_P^2}{2c_{\parallel}^2(z)} \right]^{-1}, \quad (157)$$

where

$$k_s \equiv \sqrt{k_{s\perp}^2 + k_{s\parallel}^2}, \quad \mu_k \equiv k_{s\parallel}/k_s, \quad \eta \equiv c_{\parallel}/c_{\perp}. \quad (158)$$

Figure 16 shows anisotropic power spectra $P^{(\text{CRD})}(k_s, \mu_k; z = 2.2)$. As specific examples, we consider SCDM, LCDM and OCDM models, which have $(\Omega_m, \Omega_\Lambda, h, \sigma_8) = (1.0, 0.0, 0.5, 0.6)$, $(0.3, 0.7, 0.7, 1.0)$, and $(0.3, 0.0, 0.7, 1.0)$, respectively. Clearly the linear theory predictions ($\sigma_P = 0$; top panels) are quite different from the results of N-body simulations (bottom panels), indicating the importance of the nonlinear velocity effects (σ_P computed according to ref. [57]; middle panels).

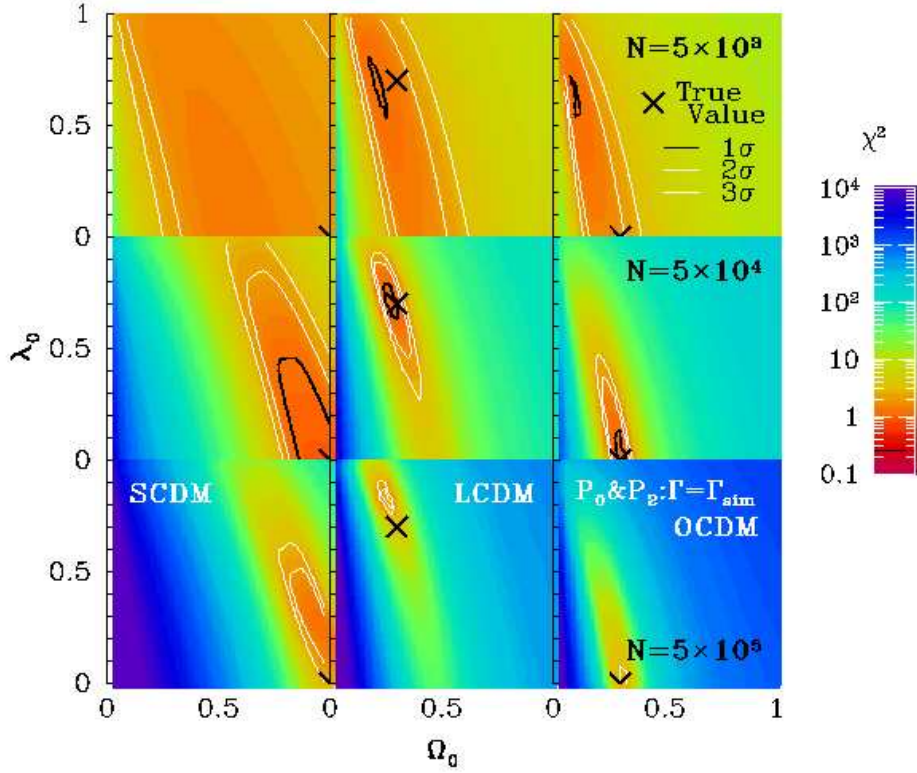


Figure 17: The confidence contours on Ω_m - Ω_Λ plane from the χ^2 -analysis of the monopole and quadrupole moments of the power spectrum in the cosmological redshift space at $z = 2.2$. We randomly selected $N = 5 \times 10^3$ (upper panels), $N = 5 \times 10^4$ (middle panels), and $N = 5 \times 10^5$ (lower panels) particles from N-body simulation. The value of σ_8 is adopted from the cluster abundance. (ref. [46])

Next we decompose the power spectrum into harmonics:

$$P(k, \mu_k; z) = \sum_{l:\text{even}} P_l(k) L_l(\mu_k), \quad P_l(k; z) \equiv \frac{2l+1}{2} \int_{-1}^1 d\mu_k P(k, \mu_k; z) L_l(\mu_k) \quad (159)$$

where $L_l(\mu_k)$ are the l -th order Legendre polynomials. Similarly, the two-point correlation function is decomposed as

$$\xi(x, \mu_x; z) = \sum_{l:\text{even}} \xi_l(x) L_l(\mu_x), \quad \xi_l(x; z) \equiv \frac{2l+1}{2} \int_{-1}^1 d\mu_x \xi(x, \mu_x; z) L_l(\mu_x) \quad (160)$$

using the direction cosine, μ_x , between the separation vector and the line-of-sight. The above multipole moments satisfy the following relations:

$$\xi_l(x; z) = \frac{1}{2\pi^2 i^l} \int_0^\infty P_l(k; z) j_l(kx) k^2 dk, \quad (161)$$

$$P_l(k; z) = 4\pi i^l \int_0^\infty \xi_l(x; z) j_l(kx) x^2 dx, \quad (162)$$

with $j_l(kx)$ being spherical Bessel functions. Substituting $P^{(\text{CRD})}(k_s, \mu_k; z)$ in equation (159) yields $P_l^{(\text{CRD})}(k_s; z)$, and then $\xi^{(\text{CRD})}(\mathbf{x}_s; z)$ can be computed from equation (161).

Comparison of the monopoles and quadrupoles from simulations and model predictions exhibits how the results are sensitive to the cosmological parameters, which in turn may put potentially useful constraints on $(\Omega_m, \Omega_\Lambda)$. Figure 17 indicates the feasibility, which interestingly results in a constraint fairly orthogonal to that from the Supernovae Ia Hubble diagram.

5.4 Two-Point Clustering Statistics on a Light-Cone in Cosmological Redshift Space

In order to explore the relation between the two-point statistics on a constant-time hypersurface in real space and that on a light-cone hypersurface in cosmological redshift space, we simply consider the case of the deterministic, linear, and scale-independent bias:

$$\delta(\mathbf{x}, z) = b(z) \delta_m(\mathbf{x}, z). \quad (163)$$

In what follows, we explicitly use the subscript, mass, to indicate the quantities related to the mass density field, and those without the subscript correspond to objects satisfying equation (163).

Using equation (157), the two-point correlation function in the cosmological redshift space, $\xi^{(\text{CRD})}(x_{s\perp}, x_{s\parallel}; z)$, is computed as

$$\xi^{(\text{CRD})}(\mathbf{x}_s; z) = \frac{1}{(2\pi)^3} \int P^{(\text{CRD})}(\mathbf{k}_s; z) \exp(-i\mathbf{k}_s \cdot \mathbf{x}_s) d^3k_s$$

$$\begin{aligned}
&= \frac{1}{(2\pi)^3} \int P^{(S)}(\mathbf{k}; z) \exp(-i\mathbf{k} \cdot \mathbf{x}) d^3k \\
&= \xi^{(S)}(c_\perp x_{s\perp}, c_\parallel x_{s\parallel}; z),
\end{aligned} \tag{164}$$

where $\xi^{(S)}(x_\perp, x_\parallel; z)$ is the redshift-space correlation function defined through equation (131).

Since $P_l^{(\text{CRD})}(k_s; z)$ and $\xi_l^{(\text{CRD})}(x_s; z)$ are defined in redshift space, the proper weight should be

$$d^3s^{(\text{CRD})} [\phi(z)n_0^{\text{CRD}}(z)]^2 = d^3x [\phi(z)n_0^{\text{com}}(z)]^2 c_\perp(z)^2 c_\parallel(z), \tag{165}$$

where $n_0^{\text{CRD}}(z)$ and $n_0^{\text{com}}(z)$ denote the number densities of the objects in cosmological redshift space and comoving space, respectively, and $\phi(z)$ is the selection function determined by the observational target selection and the luminosity function of the objects. Then, the final expressions [84] reduce to

$$P_l^{(\text{LC,CRD})}(k_s) = \frac{\int_{z_{\min}}^{z_{\max}} dz \frac{dV_c}{dz} [\phi(z)n_0^{\text{com}}(z)]^2 c_\perp(z)^2 c_\parallel(z) P_l^{(\text{CRD})}(k_s; z)}{\int_{z_{\min}}^{z_{\max}} dz \frac{dV_c}{dz} [\phi(z)n_0^{\text{com}}(z)]^2 c_\perp(z)^2 c_\parallel(z)} \tag{166}$$

$$\xi_l^{(\text{LC,CRD})}(x_s) = \frac{\int_{z_{\min}}^{z_{\max}} dz \frac{dV_c}{dz} [\phi(z)n_0^{\text{com}}(z)]^2 c_\perp(z)^2 c_\parallel(z) \xi_l^{\text{CRD}}(x_s; z)}{\int_{z_{\min}}^{z_{\max}} dz \frac{dV_c}{dz} [\phi(z)n_0^{\text{com}}(z)]^2 c_\perp(z)^2 c_\parallel(z)}, \tag{167}$$

where z_{\min} and z_{\max} denote the redshift range of the survey, $dV_c/dz = d^2_V(z)/H(z)$ is the comoving volume element per unit solid angle.

Note that k_s and x_s , defined in $P_l^{(\text{CRD})}(k_s; z)$ and $\xi_l^{\text{CRD}}(x_s; z)$, are related to their comoving counterparts at z through equations (158) and (154), while those in $P_l^{(\text{LC,CRD})}(k_s)$ and $\xi_l^{(\text{LC,CRD})}(x_s)$ are not specifically related to any comoving wavenumber and separation. Rather, they correspond to the quantities averaged over the range of z satisfying the observable conditions $x_s = (c/H_0)\sqrt{\delta z^2 + z^2 \delta \theta^2}$ and $k_s = 2\pi/x_s$.

Let us show specific examples of the two-point clustering statistics on a light-cone in cosmological redshift space. We consider SCDM and LCDM models, and take into account the selection functions relevant to the upcoming SDSS spectroscopic samples of galaxies and quasars by adopting the B -band limiting magnitudes of 19 and 20, respectively.

Figure 18 compares the predictions for the angle-averaged (monopole) power spectra under various approximations. The upper and lower panels adopt the selection functions appropriate for galaxies in $0 < z < z_{\max} = 0.2$ and QSOs in $0 < z < z_{\max} = 5$, respectively. The left and right panels present the results in SCDM and LCDM models. For simplicity we adopt a scale-independent linear

bias model [23]:

$$b(z) = 1 + \frac{1}{D(z)}[b(k, z=0) - 1], \quad (168)$$

with $b(k, z=0) = 1$ and 1.5 for galaxies and quasars, respectively.

The upper and lower panels correspond to magnitude-limited samples of galaxies ($B < 19$ in $0 < z < z_{\max} = 0.2$; no bias model) and QSOs ($B < 20$ in $0 < z < z_{\max} = 5$; Fry's linear bias model), respectively. We present the results normalized by the real-space power spectrum in linear theory $P^{(R,lin)}(k; z)$ [5], and $P_0^{(S)}(k; z=0)$, $P_0^{(S)}(k; z=z_{\max})$, $P_0^{(CRD)}(k_s; z=z_{\max})$, and $P_0^{(LC,CRD)}(k_s)$ are computed using the nonlinear power spectrum [65].

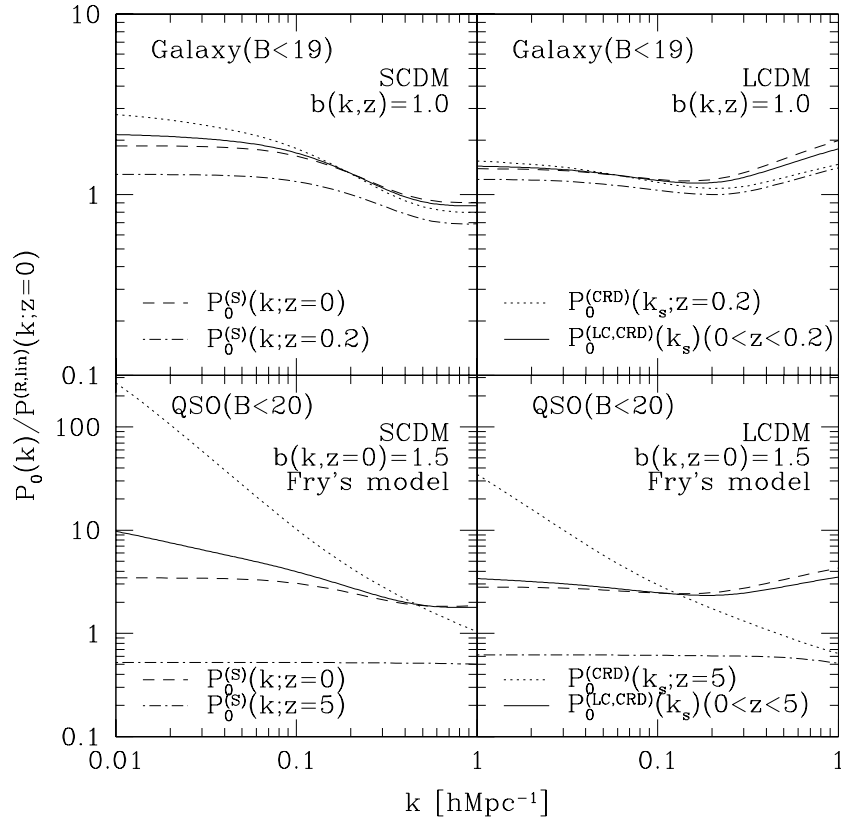


Figure 18: Light-cone and cosmological redshift-space distortion effects on angle-averaged power spectra. (ref. [84])

Consider first the results for the galaxy sample (upper panels). On linear

scales ($k < 0.1h \text{ Mpc}^{-1}$), $P_0^{(S)}(k; z = 0)$ plotted in dashed lines is enhanced relative to that in real space, mainly due to a linear redshift-space distortion [the Kaiser factor in equation (131)]. For nonlinear scales, the nonlinear gravitational evolution increases the power spectrum in real space, while the finger-of-god effect suppresses that in redshift space. Thus, the net result is sensitive to the shape and the amplitude of the fluctuation spectrum, itself; in the LCDM model that we adopted, the nonlinear gravitational growth in real space is stronger than the suppression due to the finger-of-god effect. Thus, $P_0^{(S)}(k; z = 0)$ becomes larger than its real-space counterpart in linear theory. In the SCDM model, however, this is opposite and $P_0^{(S)}(k; z = 0)$ becomes smaller.

The power spectra at $z = 0.2$ (dash-dotted lines) are smaller than those at $z = 0$ by the corresponding growth factor of the fluctuations, and one might expect that the amplitude of the power spectra on the light-cone (solid lines) would be in-between the two. While this is correct, if we use the comoving wavenumber, the actual observation on the light-cone in the cosmological redshift space should be expressed in terms of k_s [equation (158)]. If we plot the power spectra at $z = 0.2$ taking into account the geometrical distortion, $P_0^{(\text{CRD})}(k_s; z = 0.2)$ in the dotted lines becomes significantly larger than $P_0^{(S)}(k; z = 0.2)$. Therefore, $P_0^{(\text{LC,CRD})}(k_s)$ should take a value between those of $P_0^{(\text{CRD})}(k_s; z = 0) = P_0^{(S)}(k; z = 0)$ and $P_0^{(\text{CRD})}(k_s; z = 0.2)$. This explains the qualitative features shown in the upper panels of Figure 18. As a result, both the cosmological redshift-space distortion and the light-cone effect substantially change the predicted shape and amplitude of the power spectra, even for the galaxy sample [60]. The results for the QSO sample can be basically understood in a similar manner, except that the evolution of the bias makes a significant difference, since the sample extends to much higher redshifts.

Figure 19 shows the results for the angle-averaged (monopole) two-point correlation functions, exactly corresponding to those in Figure 18. The results in this figure can also be understood by an analogy of those presented in Figure 18 at $k \sim 2\pi/x$. Unlike the power spectra, however, two-point correlation functions are not positive definite. The funny features in Figure 19 on scales larger than $30h^{-1}\text{Mpc}$ ($100h^{-1}\text{Mpc}$) in SCDM (LCDM) originate from the fact that $\xi^{(\text{R,lin})}(x, z = 0)$ becomes negative there.

In fact, since the resulting predictions are sensitive to the bias, which is unlikely to quantitatively be specified by theory, the present methodology will find two completely different applications. For relatively shallower catalogues, like galaxy samples, the evolution of bias is not supposed to be so strong. Thus, one may estimate the cosmological parameters from the observed degree of the redshift distortion, as has been conducted conventionally. Most importantly, we can correct for the systematics due to the light-cone and geometrical distortion effects, which affect the estimate of the parameters by $\sim 10\%$. Alternatively, for deeper catalogues like high-redshift quasar samples, one can extract information on the object-dependent bias only by correcting the observed data on the basis

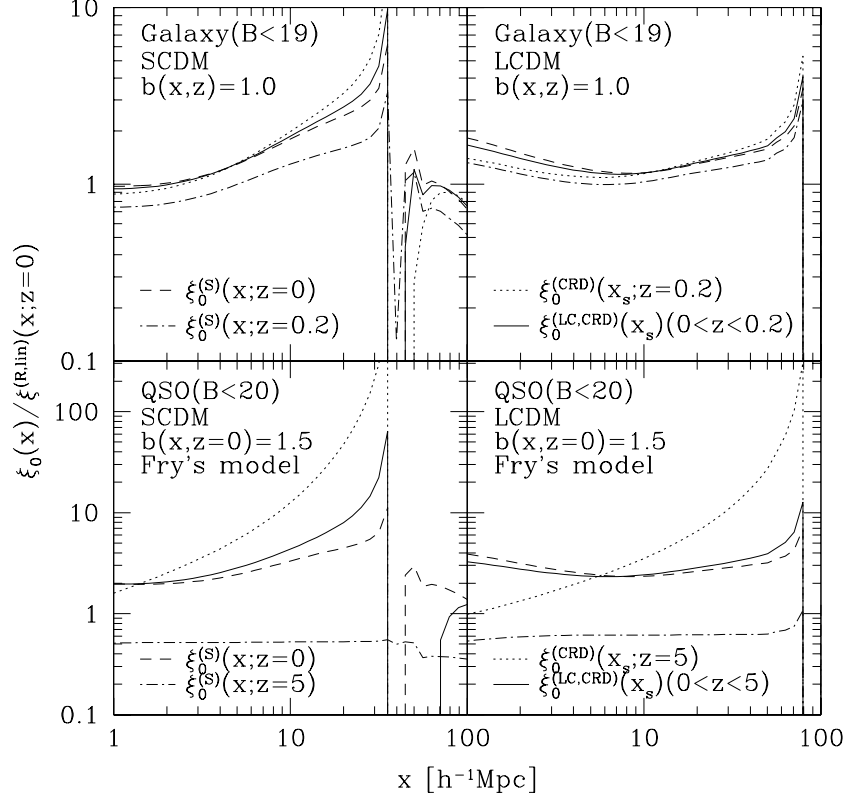


Figure 19: Same as Figure 18 on angle-averaged two-point correlation functions. (ref. [84])

of our formulae.

In a sense, the former approach uses the light-cone and geometrical distortion effects as real cosmological signals, while the latter regards them as inevitable, but physically removable, noise. In both cases, the present methodology is important in properly interpreting the observations of the universe at high redshifts.

6 Recent results from 2dF and SDSS

6.1 The latest galaxy redshift surveys

Redshifts surveys in the 1980s and the 1990s (e.g the CfA, IRAS and Las Campanas surveys) measured thousands to tens of thousands galaxy redshifts. Multifibre technology now allows us to measure redshifts of millions of galaxies. Below we summarize briefly the properties of the main new surveys, 2dFGRS, SDSS, 6dF, VIRMOS, DEEP2 and we discuss key results from 2dFGRS and SDSS. Further analysis of these surveys is currently underway.

6.1.1 The 2dF galaxy redshift survey

The Anglo-Australian 2 degree Field Galaxy Redshift (2dFGRS)¹ has recently been completed with redshifts for 230,000 galaxies selected from the APM catalogue December 2002) down to an extinction corrected magnitude limit of $b_J < 19.45$. The main survey regions are two declination strips, in the northern and southern Galactic hemispheres, and also 100 random fields, covering in total about 1800 deg^2 (Figs. 20 and 21). The median redshift of the 2dFGRS is $\bar{z} \sim 0.1$ (see refs. [11, 67] for reviews).

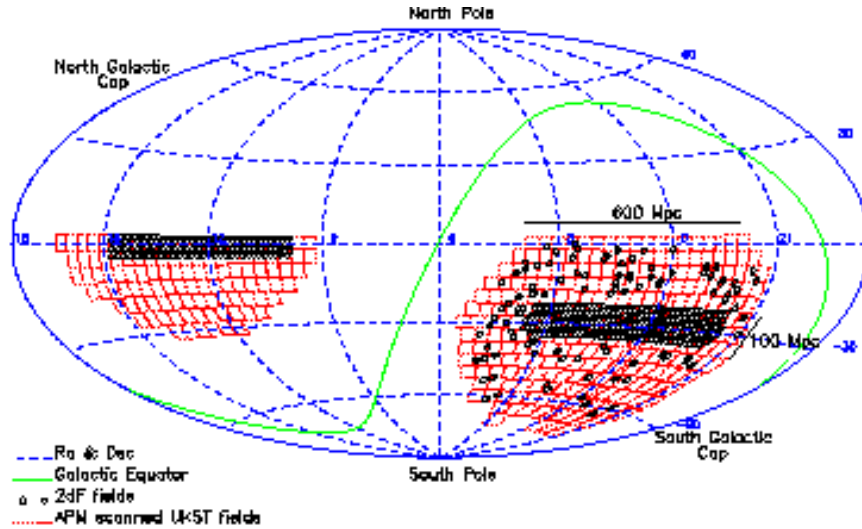


Figure 20: The 2dFGRS fields (small circles) superimposed on the APM catalogue area (dotted outlines of Sky Survey plates).

¹<http://www.mso.anu.edu.au/2dFGRS/>

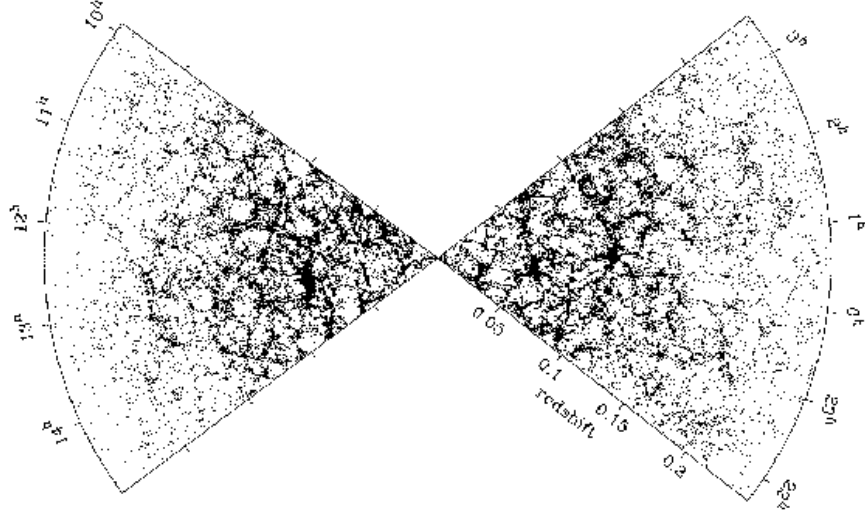


Figure 21: The distribution of 63,000 2dFGRS galaxies in the NGP (*Left*) and SGP (*Right*) strips.

6.1.2 The SDSS galaxy redshift survey

The SDSS (Sloan Digital Sky Survey) is a US-Japan-Germany joint project to image a quarter of the Celestial Sphere at high Galactic latitude as well as to obtain spectra of galaxies and quasars from the imaging data². The dedicated 2.5 meter telescope at Apache Point Observatory is equipped with a multi-CCD camera with five broad bands centered at 3561, 4676, 6176, 7494, and 8873 Å. For further details of SDSS, see refs. [97, 80]

The latest map of the SDSS galaxy distribution, together with a typical slice, are shown in Figures 22, and 23 [31]. The three-dimensional map centered on us in equatorial coordinate system is shown Figure 22. Redshift slices of galaxies centered around the equatorial plane with various redshift limits and thicknesses of planes are shown in Figure 23: $z < 0.05$ with thickness of $10h^{-1}\text{Mpc}$ centered around the equatorial plane in the upper-left panel; $z < 0.1$ with thickness of $15h^{-1}\text{Mpc}$ in the upper-right panel; $z < 0.2$ with thickness of $20h^{-1}\text{Mpc}$ in the lower panel.

²<http://www.sdss.org/>

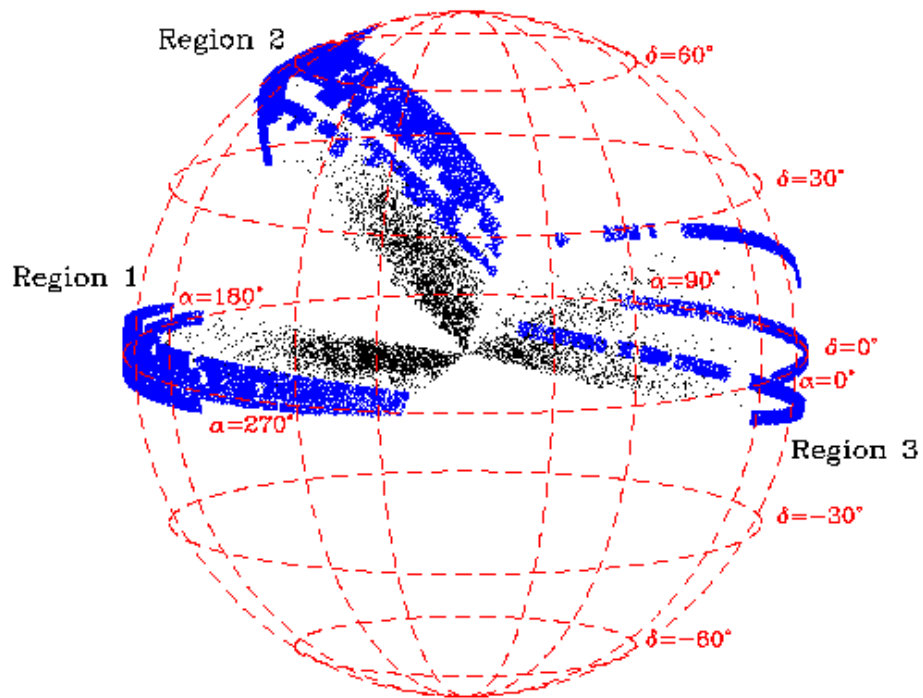


Figure 22: 3D redshift-space map centered on us, and its projection on the celestial sphere of SDSS galaxy subset, including the three main regions. (ref. [31])

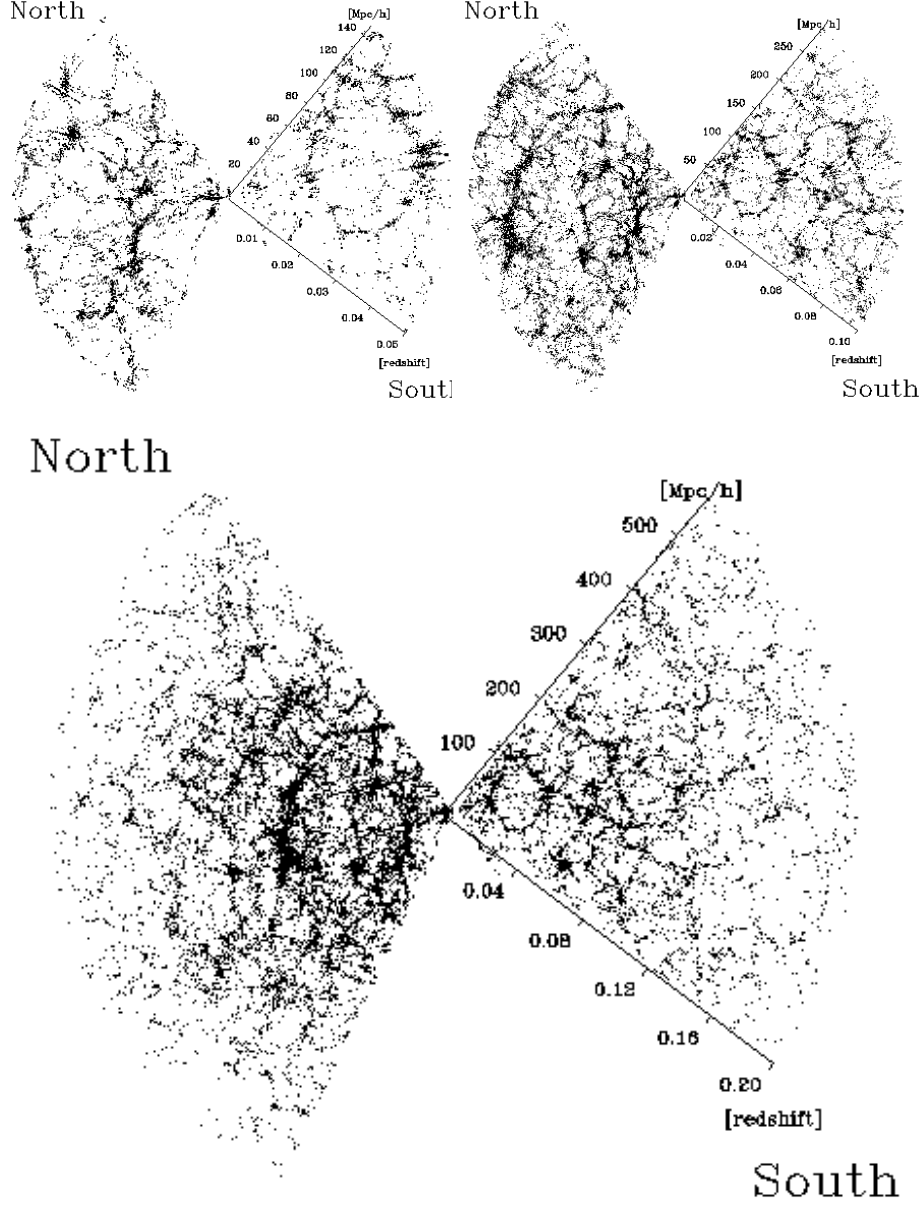


Figure 23: Redshift slices of SDSS galaxy data around the equatorial plane. The redshift limits and the thickness of the planes are: *Upper-left* $z < 0.05$, $10h^{-1}\text{Mpc}$; *Upper-right* $z < 0.1$, $15h^{-1}\text{Mpc}$; *Lower* $z < 0.2$, $20h^{-1}\text{Mpc}$. The size of points has been adjusted. Note that the data for the Southern part are sparser than those for the Northern part, especially for thick slices. (ref. [31])

6.1.3 The 6dF galaxy redshift survey

The 6dF (6-degree Field)³ is a survey of redshifts and peculiar velocities of galaxies selected primarily in the Near Infrared from new 2MASS (Two Micron All Sky Survey) catalogue⁴. One goal is to measure redshifts of more than 170,000 galaxies over nearly the whole Southern sky. Another exciting aim of the survey is to measure peculiar velocities (using 2MASS photometry and 6dF velocity dispersions) of about 15,000 galaxies out to $150h^{-1}\text{Mpc}$. The high quality data of this survey could revive peculiar velocities as a cosmological probe (which was very popular about 10-15 years ago). Observations have so far obtained nearly 40,000 redshifts and completion is expected in 2005.

6.1.4 The DEEP galaxy redshift survey

The DEEP survey is a two-phased project using the Keck telescopes to study the properties and distribution of high redshift galaxies⁵. Phase 1 used the LRIS spectrograph to study a sample of ~ 1000 galaxies to a limit of $I=24.5$. Phase 2 of the DEEP project will use the new DEIMOS spectrograph to obtain spectra of $\sim 65,000$ faint galaxies with redshifts $z \sim 1$. The scientific goals are to study the evolution of properties of galaxies and the evolution of the clustering of galaxies compared to samples at low redshift. The survey is designed to have the fidelity of local redshift surveys such as the LCRS survey, and to be complementary to ongoing large redshift surveys such as the SDSS project and the 2dF survey. The DEIMOS/DEEP or DEEP2 survey will be executed with resolution $R\ 4000$, and we therefore expect to measure linewidths and rotation curves for a substantial fraction of the target galaxies. DEEP2 will thus also be complementary to the VLT/VIRMOS project, which will survey more galaxies in a larger region of the sky but with much lower spectral resolution and with fewer objects at high redshift.

6.1.5 The VIRMOS galaxy redshift survey

The on-going Franco-Italian VIRMOS project⁶ has delivered the VIMOS spectrograph for the European Southern Observatory Very Large Telescope (ESO-VLT). VIMOS is a VISIBLE imaging Multi-Object Spectrograph with outstanding multiplex capabilities: with 10 arcsec slits, spectra can be taken of 600 objects simultaneously. In integral field mode, a 6400-fibre Integral Field Unit (IFU) provides spectroscopy for all objects covering a $54 \times 54 \text{ arcsec}^2$ area. VIMOS therefore provides unsurpassed efficiency for large surveys. The VIRMOS project consists of: Construction of VIMOS, and a Mask Manufacturing Unit for the ESO-VLT. The VIRMOS-VLT Deep Survey (VVDS), a comprehensive

³<http://www.mso.anu.edu.au/6dFGS/>

⁴<http://www.ipac.caltech.edu/2mass/>

⁵<http://deep.berkeley.edu/marc/deep/>

⁶<http://www.astrsp-mrs.fr/virmos/>

imaging and redshift survey of the deep universe based on more than 150,000 redshifts in four 4 sq.-degree fields.

6.2 Cosmological parameters from 2dFGRS

6.2.1 The Power spectrum of 2dF Galaxies on large scales

An initial estimate of the convolved, redshift-space power spectrum of the 2dFGRS was determined by Percival et al. (2002) [72] for a sample of 160,000 redshifts. On scales $0.02h/\text{Mpc} < k < 0.15h/\text{Mpc}$, the data are fairly robust and the shape of the power spectrum is not significantly affected by redshift-space distortion or non-linear effects, while its overall amplitude is increased due to the linear redshift-space distortion effect (see §5).

If one fits the Λ -CDM model predictions to the 2dFGRS power spectrum (Fig. 24) over the above range in k , one can constrain the cosmological parameters. For instance, assuming a Gaussian prior on the Hubble constant $h = 0.7 \pm 0.07$ (from ref. [22]), Percival et al. (2002) [72] obtained the 68 percent confidence limits on the shape parameter $\Omega_m h = 0.20 \pm 0.03$, and the baryon fraction $\Omega_b/\Omega_m = 0.15 \pm 0.07$. For a fixed set of cosmological parameters, i.e., $n = 1$, $\Omega_m = 1 - \Omega_\Lambda = 0.3$, $\Omega_b h^2 = 0.02$ and $h = 0.70$, the rms mass fluctuation amplitude of 2dFGRS galaxies smoothed over a top-hat radius of $8h^{-1}\text{Mpc}$ in redshift space turned out to be $\sigma_{8g}^S(L_s, z_s) \approx 0.94$.

6.2.2 An upper limit on neutrino masses

The recent results of atmospheric and solar neutrino oscillations [24, 1] imply non-zero mass-squared differences of the three neutrino flavours. While these oscillation experiments do not directly determine the absolute neutrino masses, a simple assumption of the neutrino mass hierarchy suggests a lower limit on the neutrino mass density parameter, $\Omega_\nu = m_{\nu,\text{tot}} h^{-2}/(94\text{eV}) \approx 0.001$. Large scale structure data can put an upper limit on the ratio Ω_ν/Ω_m due to the neutrino ‘free streaming’ effect [33]. By comparing the 2dF galaxy power-spectrum of fluctuations with a four-component model (baryons, cold dark matter, a cosmological constant and massive neutrinos) it was estimated that $\Omega_\nu/\Omega_m < 0.13$ (95% CL), or with concordance prior of $\Omega_m = 0.3$, $\Omega_\nu < 0.04$, or an upper limit of $\sim 2\text{ eV}$ on the total neutrino mass, assuming a prior of $h \approx 0.7$ [17, 18] (see Fig. 24). In order to minimize systematic effects due to biasing and non-linear growth, the analysis was restricted to the range $0.02 < k < 0.15 h \text{ Mpc}^{-1}$. Additional cosmological data sets bring down this upper limit by a factor of two [79].

6.2.3 Combining 2dFGRS & CMB

While the CMB probes the fluctuations in matter, the galaxy redshift surveys measure the perturbations in the light distribution of particular tracer (e.g.

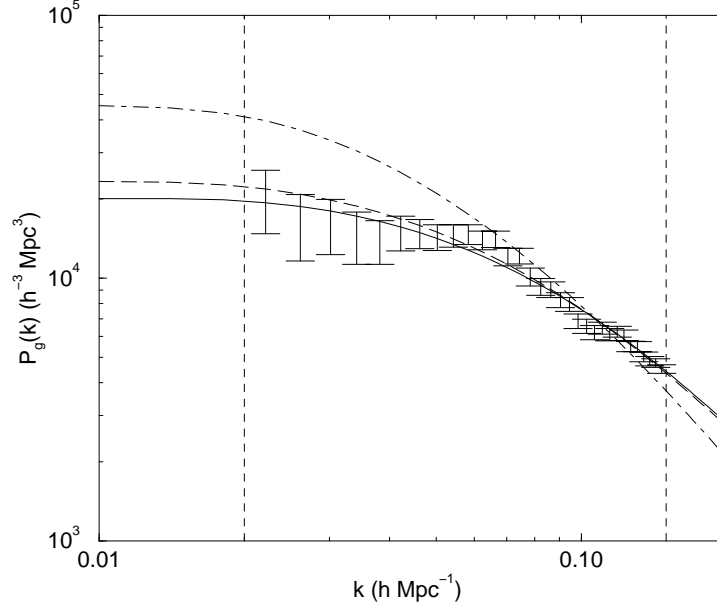


Figure 24: The power spectrum of the 2dFGRS. The points with error bars show the measured 2dFGRS power spectrum measurements in redshift space, convolved with the window function. Also plotted are linear CDM models with neutrino contribution of $\Omega_\nu = 0, 0.01$ and 0.05 (bottom to top lines) The other parameters are fixed to the concordance model. The good fit of the linear theory power spectrum at $k > 0.15 h \text{ Mpc}^{-1}$ is due to a conspiracy between the non-linear gravitational growth and the Finger-of-God smearing. (refs. [72, 17])

galaxies of certain type). Therefore, for a fixed set of cosmological parameters, a combination of the two can constrain better cosmological parameters, and it can also provide important information on the way galaxies are ‘biased’ relative to the mass fluctuations,

The CMB fluctuations are commonly represented by the spherical harmonics C_ℓ . The connection between the harmonic ℓ and k is roughly

$$\ell \simeq k \frac{2c}{H_0 \Omega_m^{0.4}} \quad (169)$$

for a spatially-flat universe. For $\Omega_m = 0.3$ the 2dFGRS range $0.02 < k < 0.15 h \text{ Mpc}^{-1}$ corresponds approximately to $200 < \ell < 1500$, which is well covered by the recent CMB experiments.

Recent CMB measurements have been used in combination with the 2dF power spectrum. Efstathiou et al. (2002) [19] showed that 2dFGRS+CMB

provide evidence for a positive Cosmological Constant $\Omega_\Lambda \sim 0.7$ (assuming $w = -1$), independently of the studies of Supernovae Ia. As explained in ref. [72], the shapes of the CMB and the 2dFGRS power spectra are insensitive to Dark Energy. The main important effect of the Dark Energy is to alter the angular diameter distance to the last scattering, and thus the position of the first acoustic peak. Indeed, the latest result from a combination of WMAP with 2dFGRS and other probes gives $h = 0.71^{+0.04}_{-0.03}$, $\Omega_b h^2 = 0.0224 \pm 0.0009$, $\Omega_m h^2 = 0.135^{+0.008}_{-0.009}$, $\sigma_8 = 0.84 \pm 0.04$, $\Omega_{\text{tot}} = 1.02 \pm 0.02$, and $w < -0.78$ (95% CL, assuming $w \geq -1$) [79].

6.2.4 Redshift-space distortion

An independent measurement of cosmological parameters on the basis of 2dFGRS comes from redshift-space distortions on scales $\lesssim 10 h^{-1} \text{Mpc}$; a correlation function $\xi(\pi, \sigma)$ in parallel and transverse pair separations π and σ . As described in §5, the distortion pattern is a combination of the coherent infall, parameterized by $\beta = \Omega_m^{0.6}/b$ and random motions modelled by an exponential velocity distribution function (eq.[133]). This methodology has been applied by many authors. For instance, Peacock et al. (2001) [66] derived $\beta(L_s = 0.17, z_s = 1.9L_*) = 0.43 \pm 0.07$ and Hawkins et al. (2003) [30] obtained $\beta(L_s = 0.15, z_s = 1.4L_*) = 0.49 \pm 0.09$ and velocity dispersion $\sigma_p = 506 \pm 52 \text{ km/sec}$. Using the full 2dF+CMB likelihood function on the (b, Ω_m) plane, Lahav et al. (2002) [42] derived a slightly larger (but consistent within the quoted error-bars) value, $\beta(L_s = 0.17, z_s = 1.9L_*) \simeq 0.48 \pm 0.06$.

6.2.5 The bi-spectrum and higher moments

It is well established that important information on the non-linear growth of structure is encoded at the high order moments, e.g. the skewness or its Fourier version, the bi-spectrum. Verde et al. [92]) computed the bi-spectrum of 2dFGRS and used it to measure the bias parameter of the galaxies. They assumed a specific quadratic biasing model:

$$\delta_g = b_1 \delta_m + \frac{1}{2} b_2 \delta_m^2.$$

By analysing 80 million triangle configurations in the wavenumber range $0.1 < k < 0.5 h/\text{Mpc}$ they found $b_1 = 1.04 \pm 0.11$ and $b_2 = -0.054 \pm 0.08$, in support of no biasing on large scale. This is a non-trivial result, as the analysis covers non-linear scales. Baugh et al. [3] and Croton et al. [14]) measured the moments of the galaxy count probability distribution function in 2dFGRG up to order $p = 6$ (order $p = 2$ is the variance, $p = 3$ is the skewness, etc.). They demonstrated the hierarchical scaling of the averaged p -point galaxy correlation functions. However, they found that the higher moments are strongly affected by the presence of two massive superclusters in the 2dFGRS volume. This poses the question of whether 2dFGRS is a 'fair sample' for high order moments.

6.3 Luminosity and spectral-type dependence of galaxy clustering

Although biasing was commonly neglected until the early 1980s, it has become evident observationally that on scales $\lesssim 10 h^{-1} \text{Mpc}$ different galaxy populations exhibit different clustering amplitudes, the so-called morphology-density relation [16]. As discussed in §4, galaxy biasing is naturally predicted from a variety of theoretical considerations as well as direct numerical simulations [37, 59, 15, 86, 87, 91]. Thus in this subsection, we summarize the extent to which the galaxy clustering is dependent on the luminosity, spectral-type and color of the galaxy sample from the 2dFGRS and SDSS.

6.3.1 2dFGRS: Clustering per luminosity and spectral type

Madgwick et al. (2003) [45] applied the Principal Component Analysis to compress each galaxy spectrum into one quantity, $\eta \approx 0.5 pc_1 + pc_2$. Qualitatively, η is an indicator of the ratio of the present to the past star formation activity of each galaxy. This allows one to divide the 2dFGRS into η -types, and to study e.g. luminosity functions and clustering per type. Norberg et al. (2002) [61] showed that, at all luminosities, early-type galaxies have a higher bias than late-type galaxies, and that the biasing parameter, defined here as the ratio of the galaxy to matter correlation function, $b \equiv \sqrt{\xi_g/\xi_m}$ varies as $b/b_* = 0.85 + 0.15L/L_*$. Figure 25 indicates that for L_* galaxies, the real space correlation function amplitude of η early-type galaxies is $\sim 50\%$ higher than that of late-type galaxies.

Figure 26 shows the redshift-space correlation function in terms of the line-of-sight and perpendicular to the line-of-sight separation $\xi(\sigma, \pi)$. The correlation function calculated from the most passively (‘red’, for which the present rate of star formation is less than 10 % of its past averaged value) and actively (‘blue’) star-forming galaxies. The clustering properties of the two samples are clearly distinct on scales $\lesssim 10 h^{-1} \text{Mpc}$. The ‘red’ galaxies display a prominent ‘finger-of-god’ effect and also have a higher overall normalization than the ‘blue’ galaxies. This is a manifestation of the well-known morphology-density relation. By fitting $\xi(\pi, \sigma)$ over the separation range $8 - 20 h^{-1} \text{Mpc}$ for each class, it was found that $\beta_{active} = 0.49 \pm 0.13$, $\beta_{passive} = 0.48 \pm 0.14$ and corresponding pairwise velocity dispersions σ_p of $416 \pm 76 \text{ km/sec}$ and $612 \pm 92 \text{ km/sec}$ [45]. At small separations, the real space clustering of passive galaxies is stronger than that of active galaxies, the slopes γ are respectively 1.93 and 1.50 (Fig. 27) and the relative bias between the two classes is a declining function of separation. On scales larger than $10 h^{-1} \text{Mpc}$ the biasing ratio is approaching unity.

Another statistic was applied recently by Wild et al. [94] and Conway et al. [13], of a joint counts-in-cells on 2dFGRS galaxies, classified by both color and spectral type. Exact linear bias is ruled out on all scales. The counts are better fitted to a bivariate lognormal distribution. On small scales there is

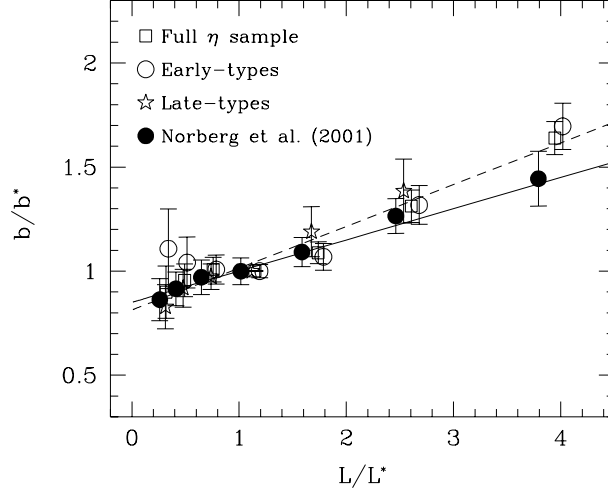


Figure 25: The variation of the galaxy biasing parameter with luminosity, relative to an L_* galaxy for the full sample and for subsamples of early and late spectra types (ref. [61]).

evidence for stochasticity. Further investigation of galaxy formation models is required to understand the origin of the stochasticity.

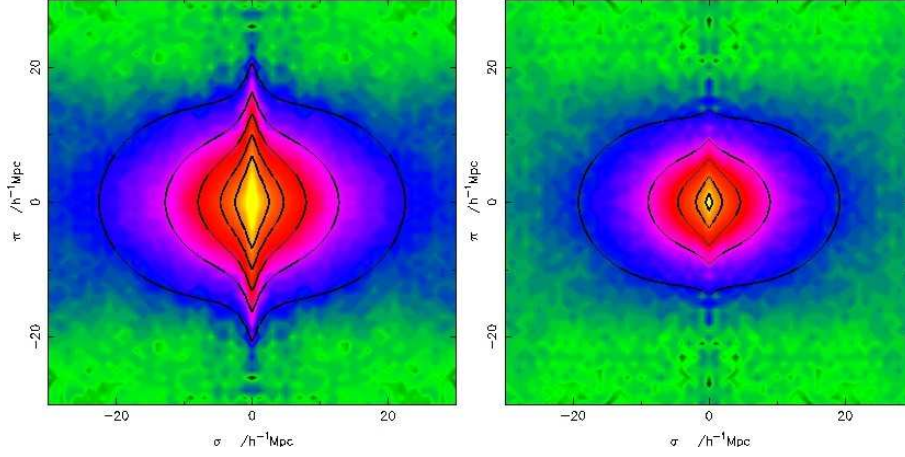


Figure 26: The two point correlation function $\xi(\sigma, \pi)$ plotted for passively (left) and actively (right) star-forming galaxies. The line contours levels show the best-fitting model (ref. [45]).

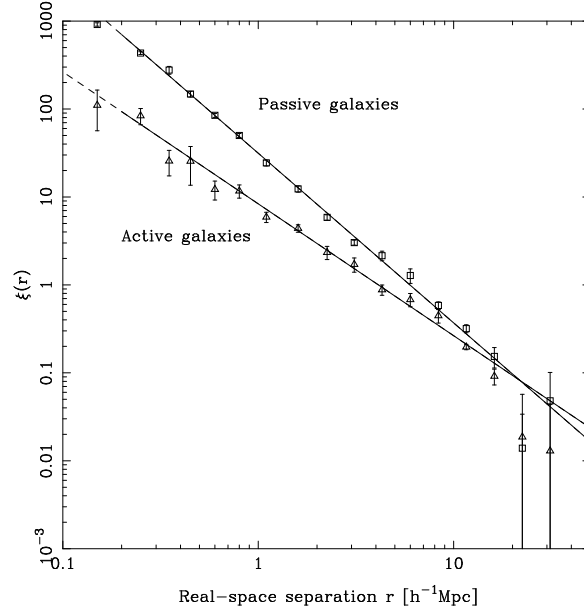


Figure 27: The correlation function for early and late spectral types. The solid lines show best-fitting models, whereas the dashes lines are extrapolations of these lines (ref. [45]).

6.3.2 SDSS: Two-point correlation functions per luminosity and color

Zehavi et al. (2002) [98] analyzed the Early Data Release (EDR) sample of the SDSS 30,000 galaxies to explore the clustering of per luminosity and color. The inferred real-space correlation function is well described by a single power-law; $\xi(r) = (r/6.1 \pm 0.2 h^{-1} \text{Mpc})^{-1.75 \pm 0.03}$, for $0.1 h^{-1} \text{Mpc} \leq r \leq 16 h^{-1} \text{Mpc}$. The galaxy pairwise velocity dispersion is $\sigma_{12} \approx 600 \pm 100 \text{ km/sec}$ for projected separations $0.15 h^{-1} \text{Mpc} \leq r_p \leq 5 h^{-1} \text{Mpc}$. When divided by color, the red galaxies exhibit a stronger and steeper real-space correlation function and a higher pairwise velocity dispersion than do the blue galaxies. In agreement with 2dFGRS there is clear evidence for scale-independent luminosity bias at $r \sim 10 h^{-1} \text{Mpc}$. Subsamples with absolute magnitude ranges centered on $M_* - 1.5$, M_* , and $M_* + 1.5$ have real-space correlation functions that are parallel power laws of slope ≈ -1.8 with correlation lengths of approximately $7.4 h^{-1} \text{Mpc}$, $6.3 h^{-1} \text{Mpc}$, and $4.7 h^{-1} \text{Mpc}$, respectively.

Figures 27 and 28 pose an interesting challenge to the theory of galaxy formation, to explain why the correlation functions per luminosity bins have similar slope, while the slope for early type galaxies is steeper than for late type.

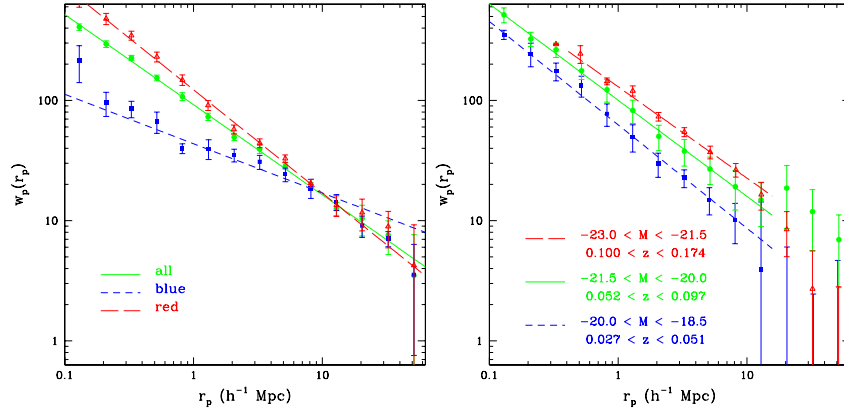


Figure 28: *Upper:* The SDSS (EDR) projected correlation function for blue (squares), red (triangles) and the full sample, with best-fitting models over the range $0.1 < r_p < 16 h^{-1} \text{Mpc}$. *Lower:* The SDSS (EDR) projected correlation function for three volume-limited samples, with absolute magnitude and redshift ranges as indicated and best-fitting power-law models. (ref. [98])

6.3.3 SDSS: Three-point correlation functions and the nonlinear biasing of galaxies per luminosity and color

Let us move next to the the three-point correlation functions (3PCF) of galaxies, which are the lowest-order unambiguous statistic to characterize non-Gaussianities due to nonlinear gravitational evolution of dark matter density fields, formation of luminous galaxies and their subsequent evolution. The determination of the 3PCF of galaxies was pioneered by Peebles & Groth (1975) and Groth & Peebles (1977) [70, 27] using the Lick and Zwicky angular catalogs of galaxies. They found that the 3PCF $\zeta(r_{12}, r_{23}, r_{31})$ obeys the *hierarchical relation*:

$$\zeta(r_{12}, r_{23}, r_{31}) = Q_r [\xi(r_{12})\xi(r_{23}) + \xi(r_{23})\xi(r_{31}) + \xi(r_{31})\xi(r_{12})] \quad (170)$$

with Q_r being a constant. The value of Q_r in real space de-projected from these angular catalogues is 1.29 ± 0.21 for $r < 3h^{-1}\text{Mpc}$. Subsequent analyses of redshift catalogs confirmed the hierarchical relation, at least approximately, but the value of Q_z (in redshift space) appears to be smaller, $0.5 \sim 1$.

As we have seen in §6.3.2, galaxy clustering is sensitive to the intrinsic properties of the galaxy samples under consideration, including their morphological types, colors, and luminosities. Nevertheless the previous analyses were not able to examine those dependences of 3PCFs because of the limited number of galaxies. Indeed Kayo et al. (2004) [39] are the first to perform the detailed analysis of 3PCFs explicitly taking account of the morphology, color and luminosity dependence. They constructed volume-limited samples from a subset of the SDSS galaxy redshift data, ‘Large-scale Structure Sample 12’. Specifically they divided each volume limited sample into color subsamples of red (blue) galaxies, which consist of 7949 (8329), 8930 (8155), and 3706 (3829) galaxies for $-22 < M_r - 5 \log h < -21$, $-21 < M_r - 5 \log h < -20$, and $-20 < M_r - 5 \log h < -19$, respectively.

Figure 29 indicates the dimensionless amplitude of the 3PCFs of SDSS galaxies in redshift space:

$$Q_z(s_{12}, s_{23}, s_{31}) \equiv \frac{\zeta(s_{12}, s_{23}, s_{31})}{\xi(s_{12})\xi(s_{23}) + \xi(s_{23})\xi(s_{31}) + \xi(s_{31})\xi(s_{12})} \quad (171)$$

for the equilateral triplets of galaxies. The overall conclusion is that Q_z is almost scale-independent and ranges between 0.5 and 1.0, and that no systematic dependence is noticeable on luminosity and color. This implies that the 3PCF itself does depend on the galaxy properties since two-point correlation functions (2PCFs) exhibit clear dependence on luminosity and color. Previous simulations and theoretical models [82, 53, 49, 85] indicate that Q decreases with scale in both real and redshift spaces. This trend is not seen in the observational results.

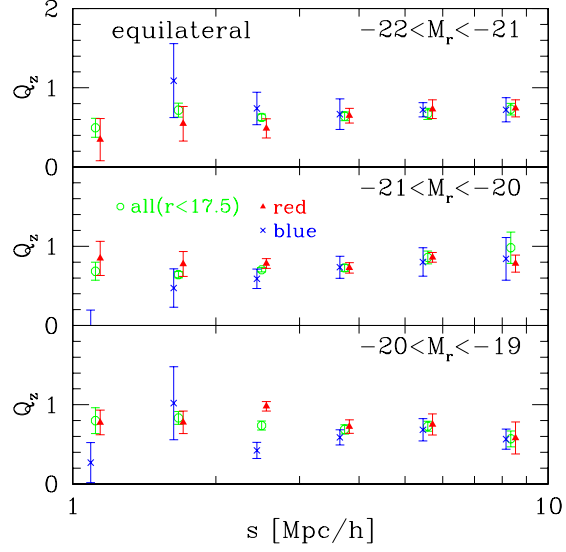


Figure 29: Dimensionless amplitude of the three-point correlation functions of SDSS galaxies in redshift space. The galaxies are classified according to their colors; all galaxies in open circles, red galaxies in solid triangles, and blue galaxies in crosses. (ref. [39]).

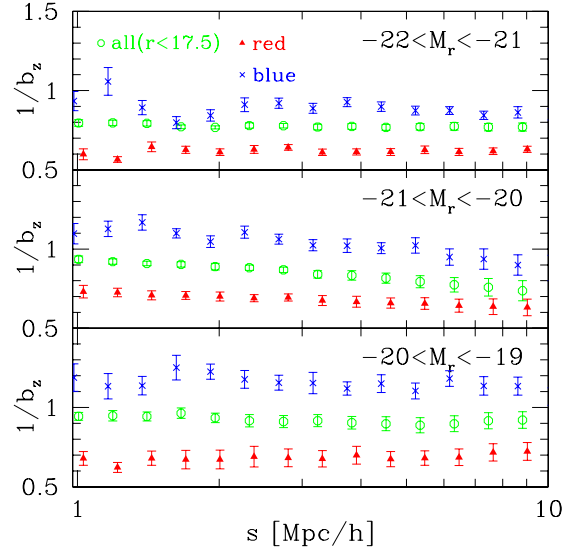


Figure 30: Same as Fig.29 but for the inverse of the biasing parameter defined through the two-point correlation functions. (ref. [39]).

In order to demonstrate the expected dependence in the current samples, they compute the biasing parameters estimated from the 2PCFs:

$$b_{z,i}(s) \equiv \sqrt{\frac{\xi_{z,i}(s)}{\xi_{z,\Lambda\text{CDM}}(s)}}, \quad (172)$$

where the index i runs over each sample of galaxies with different colors and luminosities. The predictions of the mass 2PCFs in redshift space, $\xi_{z,\Lambda\text{CDM}}(s)$, in the Λ cold dark matter model are computed following ref. [28].

As an illustrative example, consider a simple bias model in which the galaxy density field $\delta_{g,i}$ for the i -th population of galaxies is given by

$$\delta_{g,i} = b_{g,i(1)}\delta_{\text{mass}} + b_{g,i(2)}\delta_{\text{mass}}^2. \quad (173)$$

If both $b_{g,i(1)}$ and $b_{g,i(2)}$ are constant and mass density field $\delta_{\text{mass}} \ll 1$, equation (171) implies that

$$Q_{g,i} = \frac{1}{b_{g,i(1)}}Q_{\text{mass}} + \frac{b_{g,i(2)}}{b_{g,i(1)}^2}. \quad (174)$$

Thus the linear bias model ($b_{g,i(2)} = 0$) simply implies that $Q_{g,i}$ is inversely proportional to $b_{g,i(1)}$, which is plotted in Figure 30. Comparison of Figures 29 and 30 indicates that the biasing in the 3PCFs seems to compensate the difference of Q_g purely due to that in the 2PCFs.

Such behavior is unlikely to be explained by any simple model inspired by the perturbative expansion like equation (173). Rather it indeed points to a kind of regularity or universality of the clustering hierarchy behind galaxy formation and evolution processes. Thus the galaxy biasing seems much more complex than the simple deterministic and linear model. More precise measurements of 3PCFs and even higher-order statistics with future SDSS datasets would be indeed valuable to gain more specific insights into the empirical biasing model.

6.4 Topology of the universe: analysis of SDSS galaxies in terms of Minkowski functionals

All the observational results presented in the preceding subsections were restricted to the two-point statistics. As emphasized in §3, the clustering pattern of galaxies has much richer content than the two-point statistics can probe. Historically the primary goal of the topological analysis of galaxy catalogues: was to test Gaussianity of the primordial density fluctuations. Although the major role for that goal has been superseded by the CMB map analysis [41], the proper characterization of the morphology of large-scale structure beyond the two-point statistics is of fundamental importance in cosmology. In order to illustrate a possibility to explore the topology of the universe by utilizing the

new large surveys, we summarize the results of the Minkowski Functionals (MF) analysis of SDSS galaxy data [31].

In an apparent-magnitude limited catalogue of galaxies, the average number density of galaxies decreases with distance because only increasingly bright galaxies are included in the sample at larger distance. With the large redshift surveys it is possible to avoid this systematic change in both density and galaxy luminosity by construct volume-limited samples of galaxies, with cuts on both absolute-magnitude and redshift. This is in particular useful for analyses such as MF and was carried out in the analysis shown here.

Figure 31 shows the MFs as a function of ν_f defined from the volume fraction [26]:

$$f = \frac{1}{\sqrt{2\pi}} \int_{\nu_f}^{\infty} e^{-x^2/2} dx. \quad (175)$$

This is intended to map the threshold so that the volume fraction on the high-density side of the isodensity surface is identical to the volume in regions with density contrast $\delta = \nu_f \sigma$, for a Gaussian random field with r.m.s. density fluctuations σ . If the evolved density field may approximately have a good one-to-one correspondence with the initial random-Gaussian field, then this transformation removes the effect of evolution of the PDF of the density field. Under this assumption, the MFs as a function of volume fraction would be sensitive only to the topology of the isodensity contours rather than evolution with time of the density threshold assigned to a contour. While the limitations of the approximation of monotonicity in the relation between initial and evolved density fields are well recognized [40], we plot the result in this way for simplicity.

The good match between the observed MFs and the mock predictions based on the LCDM model with the initial random-Gaussianity, as illustrated in Figure 31, might be interpreted to imply that the primordial Gaussianity is confirmed. A more conservative interpretation is that, given the size of the estimated uncertainties, these data do not provide evidence for initial non-Gaussianity, i.e., the data are *consistent* with primordial Gaussianity. Unfortunately due to the statistical limitation of the current SDSS data, it is not easy to put more quantitative statement concerning the initial Gaussianity. Moreover, in order to go further and place more quantitative constraints on primordial Gaussianity with upcoming data, one needs a more precise and reliable theoretical model for the MFs which properly describes the nonlinear gravitational effect possibly as well as galaxy biasing beyond the simple mapping on the basis of the volume fraction. In fact, galaxy biasing is a major source of uncertainty for relating the observed MFs to those obtained from the mock samples for dark matter distributions. If LCDM is the correct cosmological model, the good match of the MFs for mock samples from the LCDM simulations to the observed SDSS MFs may indicate that nonlinearity in the galaxy biasing is relatively small, at least small enough that it does not significantly affect the MFs (the MFs as a function of ν_σ remain unchanged for the linear biasing).

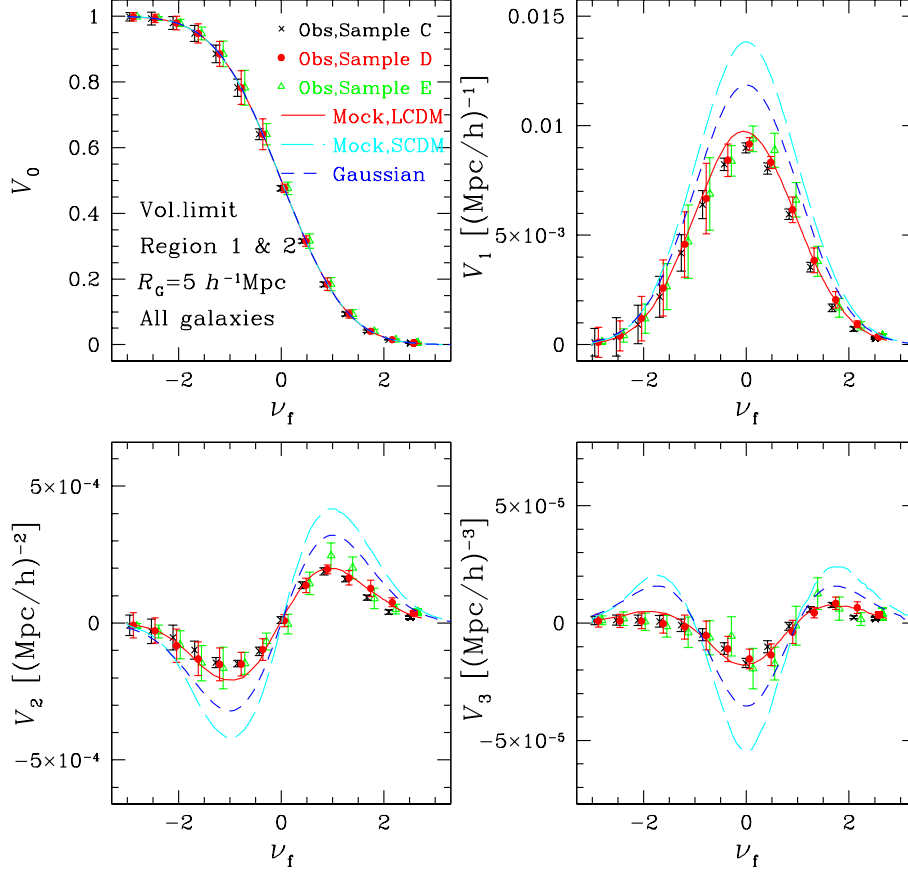


Figure 31: MFs as a function of ν_f for $R_G = 5h^{-1}\text{Mpc}$ for SDSS data. Averaged MFs of the mock samples are plotted for LCDM (solid lines) and SCDM (long dashed lines). Gaussian model predictions (eqs.[112] to [115]) are also plotted with short dashed lines. The results favor the LCDM model with random-Gaussian initial conditions. (ref. [31]).

6.5 Other statistical measures

In this section, we have presented the results on the basis of particular statistical measures including the two-point correlation functions, power spectrum, redshift distortion, and Minkowski functionals. Of course there are other useful approaches in analysing redshift surveys; the void probability function, counts-in-cells, Voronoi cells, percolation, and minimal spanning trees. Another area not covered here is optimal reconstruction of density field (e.g, using the Wiener filtering). The reader is referred to a good summary of those and other methodology in the book by Martinez & Saar (2002) [48].

Admittedly the results that we presented here are rather observational and phenomenological, and far from being well-understood theoretically. It is quite likely that when other on-going and future surveys are analysed in great detail, the nature of galaxy clustering will be revealed in a much more quantitative manner. They are supposed to act as a bridge between cosmological framework and galaxy formation operating in the universe. While the proper understanding of physics of galaxy formation is still far away, the future redshift survey data will present interesting challenges for constructing models of galaxy formation.

7 Discussion

As a classical probe, galaxy redshift surveys still remain an important tool for studying cosmology and galaxy formation. On the large scales ($> 10h^{-1}$ Mpc or so) they nicely complement the Cosmic Microwave Background, Supernovae Ia and Gravitational Lensing in quantifying in detail the cosmological model. On the small scales ($< 10h^{-1}$ Mpc) the clustering patterns of different galaxy types (defined by structural or spectral properties) provide important constraints on models of biased galaxy formation.

The redshift surveys mainly constrain Ω_m via both redshift distortion (which also depends on biasing) and the shape of the Λ -CDM power spectrum, which depends on the primordial spectrum, the product $\Omega_m h$ and also the baryon density Ω_b . Redshift surveys at a given epoch are not sensitive to the Dark Energy (or the Cosmological Constant, in a specific case), but combined with the CMB they can constrain the cosmic Equation of State.

A good example of the importance of the redshift surveys in Cosmology is given by the recent WMAP analysis of cosmological parameters, where the estimation of certain parameters was much improved by adding the 2dF power spectrum of fluctuations ([79]) or the SDSS power spectrum ([88]). This is illustrated in Table 2 by contrasting the WMAP alone derived parameters from those derived from WMAP+SDSS ([88]). Such results are sensitive to the assumed parameter space and priors, but for simplicity we quote here the results for the simple six-parameter model. In this analysis it was assumed that the universe is flat, the fluctuations are adiabatic, no gravity waves, no running tilt of the spectral index, negligible neutrino masses and that the Dark Energy is in the form of Einstein's Cosmological constant ($w = -1$). Within the Λ -CDM model this scenario can be characterised by the six parameters given in Table 2 based on [88]. The WMAP data used are both the temperature and polarization fluctuations. It can be seen that by adding the SDSS information more than halves the WMAP-only error bars on some of the parameters. These results are in good agreement with the joint analysis WMAP+2dF ([79]).

We emphasize that these parameters were fitted assuming the Λ -CDM model. While the degree of such phenomenological successes of the Λ -CDM model is truly amazing, there are many fundamental open questions:

- (i) Both components of the model, Λ and CDM, have not been directly measured. Are they ‘real’ entities or just ‘epicycles’? Historically epicycles were actually quite useful in forcing observers to improve their measurements and theoreticians to think about better models!
- (ii) ‘The Old Cosmological Constant problem’: Why is Ω_Λ at present so small relative to what is expected from Early Universe physics?
- (iii) ‘The New Cosmological Constant problem’: Why is $\Omega_m \sim \Omega_\Lambda$ at the present-epoch? Why is $w \sim -1$? Do we need to introduce new physics

Table 2: Derived 6 cosmological parameters from WMAP alone (temperature & polarization) vs. WMAP+SDSS, from Tables 2 and 3 of Tegmark et al. (2003).[88] The quoted error bars correspond to $1-\sigma$.

Symbol	WMAP alone	WMAP+SDSS	description
Ω_Λ	$0.75^{+0.10}_{-0.10}$	$0.699^{+0.042}_{-0.045}$	dimensionless cosmological constant
$\Omega_b h^2$	$0.0245^{+0.0050}_{-0.0019}$	$0.0232^{+0.0013}_{-0.0010}$	baryon density parameter
$\Omega_m h^2$	$0.140^{+0.020}_{-0.018}$	$0.1454^{+0.0091}_{-0.0082}$	total matter density parameter
σ_8	$0.99^{+0.19}_{-0.14}$	$0.917^{+0.090}_{-0.072}$	mass fluctuation amplitude at $8h^{-1}\text{Mpc}$ sphere (linear theory)
n_s	$1.02^{+0.16}_{-0.06}$	$0.977^{+0.039}_{-0.025}$	primordial scalar spectral index at $k = 0.05/\text{Mpc}$
τ	$0.21^{+0.24}_{-0.11}$	$0.124^{+0.083}_{-0.057}$	re-ionization optical depth

or to invoke the Anthropic Principle to explain it ?

- (iv) There are still open problems in Λ -CDM on the small scales e.g. galaxy profiles and satellites.
- (v) Could other (yet unknown) models fit the data equally well ?
- (vi) Where does the field go from here ? Should the activity focus on refinement of the cosmological parameters within Λ -CDM, or on introducing entirely new paradigms ?

Even if Λ CDM turns out to be the correct model, it is not yet the “end of cosmology”. Beyond the ‘zero-th order’ task of finding the cosmological parameters of the FRW model, we would like to understand the non-linear growth of mass density fluctuations and then the formation and evolution of luminous objects. The wealth of data of galaxy images and spectra in the new surveys calls for the development of more detailed models of galaxy formation. This is important so the comparison of the measurements (e.g. correlation function per spectral type or colour) and the models could be done on equal footing, with the goal of constraining scenarios of galaxy formation. There is also room for new statistical methods to quantify the ‘cosmic web’ of filaments, clusters of voids, for effective comparison with the simulations. It may well be that in the future the cosmological parameters will be fixed by the CMB, SN Ia and other probes. Then, for fixed cosmological parameters, one may use redshift surveys primarily to study galaxy biasing and evolution with cosmic epoch.

8 Acknowledgements

We thank Joachim Wambsganss and Bernard Schutz for inviting us to write the present review article. O.L. thanks members of the 2dFGRS team and the Leverhulme Quantitative Cosmology group for helpful discussions. Y.S. thanks all his students and collaborators for over many years, in particular, Thomas Buchert, Takashi Hamana, Chiaki Hikage, Yipeng Jing, Issha Kayo, Hiromitsu Magira, Takahiko Matsubara, Hiroaki Nishioka, Jens Schmalzing, Atsushi Taruya, Kazuhiro Yamamoto, and Kohji Yoshikawa among others, for enjoyable and fruitful collaborations whose results form indeed the important elements in this review. Y.S. is also grateful for the hospitality at Institute of Astronomy, University of Cambridge, where the most of present review was put together and written up for completion. O.L. acknowledges a PPARC Senior Research Fellowship. We also thank Idit Zehavi for permitting us to use Figure 28.

Numerical simulations were carried out at ADAC (the Astronomical Data Analysis Center) of the National Astronomical Observatory, Japan (project ID: mys02, yys08a). This research was also supported in part by the Grants-in-Aid from Monbu-Kagakusho and Japan Society of Promotion of Science.

References

- [1] Ahmad, Q. R. *et al.*, *Phys. Rev. Lett.*, **87**, 071301 (2001)
- [2] Alcock, C., & Paczynski, B., *Nature*, **281**, 358 (1979)
- [3] Baugh, C.M. *et al.*, astro-ph/0401405 (2004)
- [4] Ballinger, W.E., Peacock, J.A., & Heavens, A.F., *MNRAS*, **282**, 877 (1996)
- [5] Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S., *Astrophys. J.*, **304**, 15 (1986)
- [6] Bennett, C.L., *et al.*, *Astrophys. J. Suppl.*, **148**, 1–27 (2003)
- [7] Boyle, B. J., Shanks, T., Croom, S. M., Smith, R. J., Miller, L., Loaring, N., & Heymans, C., *MNRAS*, **317**, 1014 (2000)
- [8] Bullock J. S., Wechsler, R. H., Somerville R. S., *MNRAS*, **329**, 346 (2002)
- [9] Coles, P., & Jones, B., *MNRAS*, **248**, 1 (1991)
- [10] Coles, P. & Lucchin, F., *Cosmology - The Origin and Evolution of Cosmic Structure*, 2nd edition 2002, John Wiley & Sons, Chichester

- [11] Colless, M. to appear in Carnegie Observatories Astrophysics Series, Vol.2: Measuring and Modeling the Universe, ed. W.L. Freedman (Cambridge: Cambridge University Press) (2003) astro-ph/0305051
- [12] Cooray, A. & Sheth, R., Phys.Rep., **372**, 1–129 (2002)
- [13] Conway, D.J. et al., astro-ph/0404276 (2004)
- [14] Croton, D.J. et al., astro-ph/0401434 (2004)
- [15] Dekel, A. & Lahav, O., Astrophys. J., **520**, 24 (1999)
- [16] Dressler, A., Astrophys. J., **236**, 351 (1980)
- [17] Elgaroy, O. et al., Phys Rev L, **89**, 061301 (2002)
- [18] Elgaroy, O., & Lahav, O., Journal of Cosmology and Astroparticle Physics, **4**, 004 (2003)
- [19] Efstathiou G P *et al.*, MNRAS, **330**, L29 (2002)
- [20] Ehlers, J., Geren, P. & Sachs, R.K., J Math Phys, **9(9)**, 1344 (1968)
- [21] Fan, Z. and Bardeen, J.M., Phys.Rev.D, **51**, 6714 (1995)
- [22] Freedman, W.L. et al., Astrophys. J., **553**, 47 (2001)
- [23] Fry J.N., Astrophys. J., **461**, L65 (1996)
- [24] Fukuda, S. *et al.*, *Phys. Rev. Lett.*, **85**, 3999 (2000)
- [25] Gaztañaga, E., & Yokoyama, J., Astrophys. J., **403**, 450 (1993)
- [26] Gott, J. R. III, Mellot, A. L., & Dickinson, M., Astrophys. J., **306**, 341 (1986)
- [27] Groth, E.J. & Peebles, P.J.E., Astrophys. J., **217**, 385 (1977)
- [28] Hamana, T., Colombi, S., & Suto, Y., A & A, **367** 18–26 (2001)
- [29] Hamana, T., Yoshida, N., Suto, Y., & Evrard, A.E., Astrophys. J., **561** L143–L146 (2001)
- [30] Hawkins E., et al., MNRAS, **346**, 78–96 (2003)
- [31] Hikage, C. et al., Pub.Astron.Soc.Japan, **55**, 911–931, (2003)
- [32] Hikage, C., Matsubara, T., and Suto, Y., Astrophys. J., **600**, 553–563 (2004)
- [33] Hu, W., Eisenstein, D., Tegmark, M., Phys Rev Lett, **80**, 5255 (1998)

- [34] Hubble, E. P., *Astrophys. J.*, **79**, 8 (1934)
- [35] Jing, Y. P., & Suto, Y., *Astrophys. J.*, **494**, L5 (1998)
- [36] Jing, Y.P., *Astrophys. J.*, **503**, L9 (1998)
- [37] Kaiser, N., *Astrophys. J.*, **284**, L9 (1984)
- [38] Kaiser, N., *MNRAS*, **227** 1–21 (1987)
- [39] Kayo, I., Suto, Y., Nichol, R.C., Pan, J., Szapudi, I., Connolly, A., Gardner, J., Jain, B., Kulkarni, G., Matsubara, T., Sheth, R., Szalay, A., and Brinkmann, J., *Pub.Astron.Soc.Japan*, **56** (2004) June 25 issue, in press (astro-ph/0403638)
- [40] Kayo, I., Taruya, A., & Suto, Y., *Astrophys. J.*, **561**, 22 (2001)
- [41] Komatsu, E., et al., *Astrophys. J. Suppl.*, **148**, 119–134 (2003)
- [42] Lahav, O. et al., *MNRAS*, **333**, 961 (2002)
- [43] Landy, S.D., & Szalay, A.S., *Astrophys. J.*, 412, 64 (1993)
- [44] Loveday, J., Peterson, B. A., Efstathiou, G., & Maddox, S. J., *Astrophys. J.*, **390**, 338 (1992)
- [45] Madgwick, D.S., et al., *MNRAS*, 344, 847 (2003)
- [46] Magira, H., Jing, Y.P., & Suto, Y., *Astrophys. J.*, **528** 30–50 (2000)
- [47] Magliocchetti, M., & Porciani, C., *MNRAS*, 346, 186 (2003)
- [48] Martinez, V.J. & Saar, E., *Statistics of the Galaxy Distribution*, Chapman & Hall/CRC (2002)
- [49] Matsubara, T., *Astrophys. J.*, **424**, 30 (1994)
- [50] Matsubara, T., *Astrophys. J.*, **434**, 43 (1994)
- [51] Matsubara, T., *Astrophys. J.*, **584**, 1 (2003)
- [52] Matsubara, T., *Astrophys. J.*, **591**, L79 (2003)
- [53] Matsubara, T., & Suto, Y. *Astrophys.J.*, **420**, 497 (1994)
- [54] Matsubara, T., & Suto, Y., *Astrophys. J.*, **460**, 51 (1996)
- [55] Matsubara, T., & Szalay, A.S., *Astrophys.J.*, **574**,1 (2002)
- [56] Matsubara, T., & Szalay, A.S., *Phys.Rev.Lett.*, **90**, 021302 (2003)
- [57] Mo, H.J., Jing, Y.P., & Börner, G., *MNRAS*, **286**, 979 (1997)

- [58] Mecke, K. R., Buchert, T. & Wagner, H., A&A, **288**, 697 (1994)
- [59] Mo, H., & White, S.D.M., MNRAS, **282**, 347 (1996)
- [60] Nakamura, T.T., Matsubara, T., & Suto, Y., Astrophys.J., **494**, 13–19 (1998)
- [61] Norberg P. *et al.*, MNRAS, **332**, 827 (2002)
- [62] Padmanabhan, T., “*Structure Formation in the Universe*”, (Cambridge University Press, Cambridge, 1993)
- [63] Padmanabhan, T., *Galaxies and Cosmology (Theoretical Astrophysics: Volume 3)*, Cambridge University Press (2003)
- [64] Peacock, J.A., “*Cosmological Physics*”, (Cambridge University Press, Cambridge, 1999)
- [65] Peacock, J.A. & Dodds, S.J., MNRAS, **280**, L19 (1994)
- [66] Peacock, J.A., Nature, **410**, 169(2001)
- [67] Peacock, J.A., to appear in the proceedings of the 13th Annual October Astrophysics Conference in Maryland: The Emergence of Cosmic Structure, eds. S. Holt and C. Reynolds (AIP) (2003), astro-ph/0301042
- [68] Peebles, P.J.E. , “ *Large-scale Structure of the universe*”, (Princeton University Press, Princeton, 1980)
- [69] Peebles, P.J.E. , “ *Principles of Physical Cosmology*”, (Princeton University Press, Princeton, 1993)
- [70] Peebles, P.J.E., & Groth, E.J., Astrophys. J., **196**, 1 (1975)
- [71] Pen U.L., Astrophys.J., **504**, 601 (1998)
- [72] Percival, W.J. et al., MNRAS, **337**, 1068 (2002)
- [73] Rudnicki, K., *The cosmological principles*, (Jagiellonian University, Krakow, 1995)
- [74] Saslaw, W., “*Gravitational Physics of Stellar and Galactic Systems*” (Cambridge University Press, Cambridge, 1985)
- [75] Scharf, C.A., Jahoda, K., Treyer, M., Lahav, O., Boldt, E. & Piran, T., Astrophys. J., **544**, 49 (2000)
- [76] Schmalzing, J., & Buchert, T., Astrophys. J., **482**, L1 (1997)
- [77] Seljak U., MNRAS, **318**, 203 (2000)

- [78] Somerville, R. S., Lemson, G., Sigad, Y., Dekel, A., Kauffmann, G. White, S. D. M., MNRAS, **320**, 289 (2001)
- [79] Spergel, D.N., et al., Astrophys. J. Suppl., **148**, 175–194 (2003)
- [80] Strauss, M. A., et al., Astron. J., **124**, 1810 (2002)
- [81] Sugimoto, T., & Suto, Y., Astrophys. J., **371**, 470 (1991)
- [82] Suto, Y., Prog.Theor.Phys., **90**, 1173 (1993)
- [83] Suto, Y., Magira, H., Jing, Y.P., Matsubara, T., & Yamamoto, K., Prog. Theor. Phys. Suppl., **133**, 183–210 (1999)
- [84] Suto, Y., Magira, H., & Yamamoto, K., Pub.Astron.Soc.Japan, **52**, 249–257 (2000)
- [85] Takada, M., & Jain, B., Mon.Not.R.Astron.Soc., **340**, 580 (2003)
- [86] Taruya, A. & Suto, Y., Astrophys.J., **542**, 559 (2000)
- [87] Taruya, A., Magira, H., Jing, Y.P., & Suto, Y., Pub.Astron.Soc.Japan, **53**, 155–167 (2001)
- [88] Tegmark, M., et al. Phys.Rev.D., in press (2004), astro-ph/0310723
- [89] Yamamoto, K., & Suto, Y., Astrophys. J., **517** 1–12 (1999)
- [90] Yamamoto, K., Nishioka, H., & Suto, Y., Astrophys. J., **527** 488–497 (1999)
- [91] Yoshikawa, K., Taruya, A., Jing, Y.P., & Suto, Y., Astrophysical J., **558**, 520–534 (2001)
- [92] Verde et al., MNRAS, **335**, 432 (2002),
- [93] Weinberg, S., “*Gravitation and Cosmology*”, (John Wiley & Sons, New York, 1972)
- [94] Wild, V., et al., astro-ph/0404275 (2004)
- [95] White, S. D. M., MNRAS, **186**, 145 (1979)
- [96] Wu, K.K.S., Lahav, O. & Rees, M.J., Nature, **397** , 225 (1999)
- [97] York, D. G., et al., Astron. J., **120**, 1579 (2000)
- [98] Zehavi, I. et al., Astrophys. J., **571**, 172 (2002)
- [99] Zehavi, I. et al., Astrophys.J., in press. (2003) astro-ph/0301280